University of Alberta

Evolution of the protein phosphatase type 2C family in *Caenorhabditis*

by

Paul Michael Stothard

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Molecular Biology and Genetics

Department of Biological Sciences

Edmonton, Alberta
Spring 2002

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-68626-4

Canada

University of Alberta

Library Release Form

Name of Author: Paul Michael Stothard

Title of Thesis: Evolution of the protein phosphatase type 2C family in *Caenorhabditis*

Degree: Doctor of Philosophy

Year this Degree Granted: 2002

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Paul Stothard December 17, 2001

9838 79th Avenue NW
Edmonton, Alberta
Canada T6E 1R1

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Evolution of the protein phosphatase type 2C family in *Caenorhabditis*" submitted by Paul Michael Stothard in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Molecular Biology and Genetics.

Dr. D. Pilgrim

Dr. T. Evans

Dr. J. Locke

Dr. H. McDermid

December 1, 2001

Dr. M. Pasdar

Abstract

The sexual fate of the nematode *Caenorhabditis elegans* is controlled by a complex signal transduction pathway. Previous studies suggested that some of the proteins in this pathway have evolved more quickly than non-sex-determining proteins. To investigate the causes and functional significance of rapid sex-determining protein evolution, three *C. elegans* genes encoding members of the protein phosphatase 2C (PP2C) family were compared with their orthologs from another *Caenorhabditis* species (strain CB5161). One of the genes encodes FEM-2, a sex-determining protein, while the others have no known sex-determining role. FEM-2's PP2C domain was found to be more diverged than the other PP2C domains, supporting the notion that sex-determining proteins are subjected to selective pressures that allow for or cause rapid divergence. Comparison of the positions of amino acid substitutions in FEM-2 with a solved three-dimensional structure suggests that the catalytic face of the protein is highly conserved among *C. elegans*, CB5161, and another closely related species *C. briggsae*. However, the non-conserved regions of FEM-2 cannot be said to lack functional importance, since *fem-2* transgenes from the other species were unable to rescue the germ-line defect caused by a *C. elegans fem-2* mutation. RNA-mediated interference (RNAi) was used to examine whether *fem-2* functions as a sex-determining gene in the other *Caenorhabditis* species. *fem-2(RNAi)* caused germ-line feminization in *C. elegans* hermaphrodites and males, while CB5161 worms were unaffected. In *C. briggsae*, *fem-2(RNAi)* feminized the male but not the hermaphrodite germ line. These results suggest that the role of *fem-2* in sex determination has also evolved rapidly, and may provide an example of orthologous genes regulating different processes between closely related species.

Acknowledgements

I would like to thank past and present members of the Pilgrim lab, especially Dave Pilgrim, for providing an excellent research environment. Dave Hansen, Morris Maduro, and Angela Manning were a tremendous help, always eager to share their knowledge and expertise. Wanyuan Ao, Jessica Smith, Wayne Materi, Kathy Bueble, Serene Wohlgemuth, Shawna Maguire, Chantal Denholm, Heather Lemon, Janette Berg, Tamara Checkland, and D'arcy Page were a source of good advice, clever ideas, and stimulating conversation.

Numerous researchers outside the lab have made my time at the University of Alberta enjoyable and productive, including committee members Heather McDermid and Manijeh Pasdar, neighboring professors John Bell and Ross Hodgetts, research assistant Gary Ritzel, and graduate students Gary Van Domselaar, Jane Nagel, Alan Gibbs, and Graham Banting.

I thank my friends for support and encouragement over the years, in particular Scott Warren, Ryan Johnstone, Geoff Morris, Tom Sluchinski, Sean McCaffrey, Mike Louie, and Amy Nugent. I also thank my parents Bonnie and Dave, and siblings Jenny, Mary, Neil, and Kathy. They have always supported my "search for worm genes" and have helped keep things in perspective.

Table of Contents

1. Introduction **1**

1.1 Why have sex? 1

1.2 One species two sexes 2

1.3 *C. elegans* as a model system 3

1.4 Sexual dimorphism in *C. elegans* 4

1.5 Choosing one of the sexual fates 5

1.6 The components that regulate sex determination and dosage compensation 6

1.6.1 Numerator elements in the X:A ratio 6

1.6.2 *xol-1*: the master switch gene 7

1.6.3 The *sdcs*: the branching point 8

1.7 Sex determination in the soma 9

1.7.1 *her-1* encodes a male-specific secreted protein 9

1.7.2 *tra-2* encodes a HER-1 receptor 9

1.7.3 The FEM proteins bridge the gap between TRA-2 and TRA-1 10

1.7.4 *tra-1* is the terminal regulator 11

1.8 Sex determination in the germ line 11

1.8.1 Additional genes and different forms of regulation are involved 11

1.8.2 Establishing the male fate in the hermaphrodite 12

1.8.3 Switching to the female fate in the hermaphrodite 14

1.8.4 Germ-line sex determination in the male 14

1.9 The evolution of sex determination 15

1.9.1 Sex-determining genes evolve rapidly at the sequence level 15

1.9.2 The *C. elegans* and *Drosophila* pathways are not related 16

1.10 Using FEM-2 to understand sex-determining gene and pathway evolution 17

1.10.1 A brief history of *fem-2* 18

1.10.2 The roles of other PP2Cs 19

1.11 The main objectives of this work 20

2 Materials and Methods 22

2.1 Nematode strains 22

2.2 PCR conditions and primers 22

2.3 Isolation of PP2C homologs by PCR 22

2.4 Isolation of the complete CB5161 *fem-2* cDNA 23

2.5 Northern analysis of *Cs-fem-2* expression 25

2.6 Obtaining a genomic clone of *Cs-fem-2* 25

2.7 Construction of chimeric *fem-2* genes 26

2.8 Transformation of *fem-2* animals 27

2.9 RNA-mediated interference 27

2.10 Searching for polymorphisms in *fem-2* and F25D1.1 28

2.11 Two-hybrid screen for proteins that interact with FEM-2 28

2.12 Construction and expression of *fem-2* reporter genes 29

2.13 Characterization of a human homolog of *fem-2* 30

2.14 Characterization of zebrafish pyruvate dehydrogenase phosphatase 31

2.15 Sequence analysis and manipulation 31

3 Results 34

3.1 Isolation of *Cs-fem-2* 34

3.2 Comparison of Cs-FEM-2 to Ce-FEM-2 and Cb-FEM-2 35

3.3 FEM-2's PP2C domain evolves faster than the same domain in other PP2Cs ... 35

3.4 FEM-2 has a unique catalytic region that is conserved 36

3.5 Foreign *fem-2* genes can rescue somatic feminization in *C. elegans* 37

3.6 Foreign *fem-2* genes cannot rescue germ-line feminization in *C. elegans* 38

3.7 *fem-2(RNAi)* does not feminize *C. briggsae* hermaphrodites or CB5161 males 39

3.8 Polymorphisms in *fem-2* and F25D1.1 40

3.9 Yeast two-hybrid screen for proteins that interact with FEM-2 41

3.10 Construction and expression of *fem-2* reporter genes 42

3.11 Characterization of a human homolog of *fem-2* 44

3.12 Characterization of a zebrafish pyruvate dehydrogenase phosphatase 46

4 Discussion 48

4.1 PP2Cs in CB5161 48

4.2 Cs-FEM-2 appears to be a functional PP2C 49

4.3 Rapid evolution of FEM-2 50

4.4 Portions of FEM-2 important for germ-line function may not be conserved 51

4.5 The germ-line role of *fem-2* in *C. briggsae* and CB5161 53

4.6 Emerging results from studies of *Cb-fem-1* and *Cb-fem-3* 54

4.7 How might interspecies differences between pathways arise? 55

4.8 Function of *fem-2* orthologs in more distant species 58

4.9 Future directions 61

4.9.1 Identifying FEM-2's targets and partners 61

4.9.2 Determining when and where FEM-2 is expressed 64

4.9.3 Isolating mutations in *C. briggsae* sex-determining genes 64

Tables **66**

Figures **78**

Bibliography **106**

Appendix 1 ..... **128**

Appendix 2 ..... **137**

List of Tables

Table 1. *C. elegans* sex-determining genes and proteins 66

Table 2. PP2Cs and their putative substrates 68

Table 3. Phenotypes associated with mutations in PP2C genes 69

Table 4. PCR primers used in this work 70

Table 5. *C. elegans* strains analyzed for *fem-2* and F25D1.1 polymorphisms 71

Table 6. Divergence of CB5161 and zebrafish PP2C domains 72

Table 7. Somatic phenotypes of transgenic *C. elegans fem-2(null)* animals 73

Table 8. Germ-line phenotypes of transgenic *C. elegans fem-2(b245)* animals 74

Table 9. Effects of RNAi against *fem-2* in *C. elegans*, *C. briggsae*, and CB5161 75

Table 10. Yeast two-hybrid positives obtained using FEM-2 as bait 76

Table 11. Effects of *fem-2* reporters on the fertility of *fem-2(b245)* animals 77

List of Figures

Figure 1. Overt sexual dimorphism in *C. elegans* 78

Figure 2. Development of the *C. elegans* gonad 79

Figure 3. Model of somatic sex determination 80

Figure 4. Genetic interactions that specify germ-line sex 81

Figure 5. Somatic sex determination in *Drosophila melanogaster* 82

Figure 6. Mutant alleles of *fem-2* 83

Figure 7. Three-dimensional structure of human PP2Cα 84

Figure 8. *C. elegans fem-2* primers 85

Figure 9. *C. briggsae fem-2* primers 86

Figure 10. CB5161 *fem-2* primers 87

Figure 11. Northern analysis of *Cs-fem-2* expression 88

Figure 12. Genomic structure of *Cs-fem-2* 89

Figure 13. Alignment of FEM-2 sequences with Ce-T23F11.1 and Ce-F25D1.1 90

Figure 14. Alignment of the PP2C domains isolated in this study 91

Figure 15. Phylogenetic analysis of the PP2C domains isolated in this study 92

Figure 16. Conservation of FEM-2's catalytic region 93

Figure 17. Rescue of somatic feminization in *C. elegans fem-2(null)* mutants 94

Figure 18. Results of RNAi against *fem-2* 95

Figure 19. Genomic sequence analyzed from various isolates of *C. elegans* 96

Figure 20. Site of reporter insertion in FEM-2 97

Figure 21. Expression of a FEM-2::GFP fusion in N2 animals 98

Figure 22. Expression of a FEM-2::HA fusion in *fem-2(b245)* animals 99

Figure 23. Alignment of FEM-2 sequences with KIAA0015 100

Figure 24. KIAA0015 cDNA and cosmid clones 101

Figure 25. Northern analysis of KIAA0015 expression 102

Figure 26. Expression of *pdp2* in adult zebrafish 103

Figure 27. Alignment of zebrafish and rat PDP sequences 104

Figure 28. Zebrafish *pdp2* expression in oocytes 105

List of Abbreviations

ABI.....abscisic acid insensitive

AC.....anchor cell

Apaf.....apoptosis protease-activating factor

bp.....base pair

CaM-kinase.....calmodulin-dependent protein kinase

CaMKPase.....Ca^{2+}/calmodulin-dependent kinase II phosphatase

Cb.....*Caenorhabditis briggsae*

cdc.....cell division cycle

cdk.....cyclin-dependent protein kinase

Ce.....*Caenorhabditis elegans*

ced.....cell death

CFTR.....cystic fibrosis transmembrane conductance regulator

CPEB.....cytoplasmic polyadenylation element binding

cR.....centiray

Cs.....CB5161

cut.....cuticlin

DM.....DSX and MAB-3

DMT.....DM domain gene expressed in testis

dpy.....dumpy

DSD.....developmental system drift

dsf.....dissatisfaction

dsx.....doublesex

DTC.....distal tip cell

egl.....egg laying

ERK.....extracellular signal-regulated kinase

F1A.....FEM-1-like protein in the apoptotic pathway

fbf.....*fem-3* binding factor

fem.....feminization

FIN.....FGF-inducible

fog.....feminization of germ line

fox.....feminizing on X

fru.....fruitless

GAL4AD.....GAL4 activation domain

GFP.....green fluorescent protein

gld.....defective in germ line development

HA.....hemagglutinin

her (*Caenorhabditis*).....hermaphrodization

her (*Drosophila*).....hermaphrodite

him.....high incidence of males

hpf.....hours post fertilization

IgG.....immunoglobulin G

ILK.....integrin-linked kinase

ILKAP.....integrin-linked kinase-associated phosphatase

ix.....intersex

K_a.....number of non-synonymous substitutions per site

KAPP.....kinase-associated protein phosphatase

kb.....kilobase

kbp.....kilobase pair

kDa.....kilodalton

K_s.....number of synonymous substitutions per site

laf.....lethal and feminized

mab.....male abnormal

MKK.....mitogen-activated protein kinase kinase

mog.....masculinization of germ line

msl.....male-specific lethal

nos.....nanos

PCR.....polymerase chain reaction

PDB.....protein data bank

PDP.....pyruvate dehydrogenase phosphatase

PP2C.....protein phosphatase type 2C

PPases.....phosphatases

ptc.....phosphatase two C

RACE.....rapid amplification of cDNA ends

RLK.....serine-threonine receptor-like kinase

RNAi.....RNA-mediated interference

rol.....roller

RT.....reverse transcriptase

SAPK.....stress-activated protein kinase

sdc.....sex and dosage compensation

SDS.....sodium dodecyl sulfate

SEK.....SAPK/ERK kinase

sex.....signal element on X

SL1.....spliced leader 1

SSC.....saline-sodium citrate buffer

tra.....transformer

UAP.....universal amplification primer

unc.....uncoordinated

UTR.....untranslated region

Wip.....wt-p53-induced phosphatase

X-gal.....5-bromo-4-chloro-3-indolyl beta-D-galactopyranoside

xol.....XO lethal

1. Introduction

1.1 Why have sex?

Sexual reproduction involves two fundamental processes: the formation of haploid cells by meiosis, and the fusion of compatible gametes to form a new diploid cell. Asexual reproduction is much simpler, with new organisms arising directly from mitotic cell divisions. The important difference between the two strategies is that the first shuffles the alleles of two parents and delivers new combinations to the progeny (through recombination and independent assortment of chromosomes), while the second produces genetic clones of a single parent. Which method of reproduction is better? There are many problems associated with sexual reproduction that might suggest that asexual strategies should be more successful. Sex may require individuals to spend large amounts of energy locating and interacting with suitable partners, species barriers may be vulnerable during the gamete fusion step of sex, and the recombination associated with meiosis prevents successful combinations of alleles from being maintained as a unit.

Despite these problems, most animals reproduce sexually, leading many biologists to speculate about the advantages of sex that outweigh the added complexity and cost. The most common view is that genetic recombination allows for more opportunities for improved fitness, by removing harmful mutations and allowing new combinations of alleles to come together. Some argue that the removal of harmful mutations is the only important outcome of recombination (Kondrashov 1988), while others believe that it is the allele shuffling (Barton and Charlesworth 1998). The ability to rapidly generate diverse genotypes by recombination may allow organisms

to resist parasites, and this feature alone has been suggested to be responsible for the maintenance of sex (Hamilton et al. 1990). Testing these hypotheses by experimentation is very difficult, and for the time being it is probably best to accept that since sex persists, it must be advantageous. As biologist Graham Bell at McGill University stated: "Something as complex, onerous, and laborious as sexuality is probably only going to be maintained if it's doing something very important."

1.2 One species two sexes

A feature of sexual reproduction is sexual dimorphism—the occurrence of animals in two distinct sexes. Sexual dimorphism is not confined to reproductive structures and the germ line. Any tissue that can influence reproductive success can become sexually specialized. For example, the peacock has large iridescent tail feathers, which he raises to form an enormous fan. The display is used to attract females, which lack the elongated feathers. Male caribou have large ornate antlers, which are used to defend mating spaces from competing males, while females have much smaller antlers. In both cases sexual differences in the brain control how the specialized structures are used.

The ability of individuals of the same species to follow two separate developmental pathways is as intriguing as the existence of sex. How does an initially sexually ambiguous embryo choose a sexual fate? How do the numerous sexually dimorphic structures differentiate? We are all aware that there are many human sexual dimorphisms, in part because our ability to detect these differences has evolved along with the differences themselves. However, insight into the human system is difficult

to obtain, so we have turned to model organisms in the hopes of better understanding how sexual dimorphism is achieved.

1.3 *C. elegans* as a model system

Sex-determining mechanisms have been extensively studied in the yeast *Saccharomyces cerevisiae*, the fruitfly *Drosophila melanogaster*, and the nematode *Caenorhabditis elegans*. Our lab uses *C. elegans* as a model because it exhibits complex behavioral and morphological specializations, yet is amenable to genetic and molecular analysis. Closely related nematodes can be studied using many of the techniques applied to *C. elegans*, making detailed interspecies comparisons possible. Although the fruitfly and genetics go hand in hand, and numerous sibling species are available for interspecies comparisons, it lacks some of the features that led to the worm being adopted well after *Drosophila* was established. Worms are transparent throughout development, allowing individual cell divisions and migrations to be observed. This feature is particularly helpful when studying the sex of the germ line, as sperm and oocytes can be easily identified within the intact gonad in living animals. Worms exist as self-fertile hermaphrodites or males, allowing heterozygous mutations to become homozygous in one generation (by self-fertilization), and mutations to be moved into new genetic backgrounds (using males). The hermaphrodite sex does not prevent *C. elegans* from serving as a model for conventional male/female species, since *C. elegans* hermaphrodites resemble females of other species, apart from their ability to produce a small number of sperm in their germ line prior to oogenesis. Gene function in the worm can be readily explored using DNA transformation and RNA-mediated interference (RNAi), in addition to the genetic approaches employed in other

systems like *Drosophila* and yeast. Finally, the generation time of the worm is about three days, and each hermaphrodite produces about 300 progeny, allowing the products of crosses, transformations, and RNAi experiments to be obtained quickly and in suitable numbers.

1.4 Sexual dimorphism in *C. elegans*

There are numerous differences between adult male and hermaphrodite worms, some of which are quite obvious: males are smaller than hermaphrodites; males have a tail that is fan-like and specialized for mating while hermaphrodites have a tail that lacks a fan; hermaphrodites have a vulva through which sperm can enter and fertilized eggs exit while males lack a vulva; and hermaphrodites have a two-armed gonad while males have a single-armed gonad (Figure 1). Less easily observed sexual dimorphisms are found in the nervous system and musculature. These are necessary for the male to locate and inseminate hermaphrodites, and for the hermaphrodite to lay eggs. The hermaphrodite intestine is also specialized, serving as a source of yolk for differentiating oocytes (Kimble and Sharrock 1983).

In any sexually reproducing metazoan the germ-line specializations are of primary importance. In *C. elegans* the germ cells are located inside a gonad comprised of somatic sheath cells (for review see Schedl 1997). Early during development differences between the male and hermaphrodite gonad arise, first in the way that the somatic cells migrate (Figure 2). In hermaphrodites two distal tip cells move in opposite directions from the mid-ventral region to form a two-armed gonad. In males these cells remain together, and another somatic cell, called the linker cell, migrates anteriorly and then posteriorly to the tail. The resulting gonad consists of a

single arm. In both sexes the gonad interior is populated by a syncytium of germ cells. The cells divide mitotically in the distal regions, and then switch to meiosis as they migrate towards the uterus in hermaphrodites, or the cloaca in males. The signal that determines where the mitosis to meiosis switch occurs comes from the distal tip cells (Kimble and White 1981), which extend tentacle-like processes over the distal germ cells (Hall et al. 1999). Each gonad arm in the hermaphrodite generates about 150 sperm and then switches to oocyte production, while the male gonad continuously produces sperm. Hermaphrodites store the initial sperm they produce and any sperm obtained through mating in the spermathecae. Oocytes are fertilized when they are forced through the spermathecae by the contractions of the surrounding sheath cells. After fertilization the eggs remain in the adjoining uterus for a short time before being deposited through the vulva.

1.5 Choosing one of the sexual fates

The investigation into how a worm chooses its sexual fate began over fifty years ago with the identification of the primary sex-determining signal, the ratio of X chromosomes to sets of autosomes (Nigon 1949). Individuals with two X chromosomes develop as hermaphrodites (X:A = 1), while those with a single X chromosome develop as males (X:A = 0.5). Experiments with polyploid strains have shown that the ratio can be interpreted quite accurately: if it is 0.75 or higher (3X:4A, 3X:3A, and 4X:4A) a hermaphrodite is produced, and if it is 0.67 or lower (2X:3A, 2X:4A, and 1X:2A) a male is produced (Madl and Herman 1979). Based on these findings, one would expect all the progeny of a self-fertilizing XX hermaphrodite to be XX. In practice the occasional male arises due to meiotic non-disjunction of the X

chromosome. If males are needed for experimentation, existing males can be crossed to hermaphrodites to produce 50% male outcross progeny, or mutant strains can be used in which meiotic non-disjunction occurs more frequently.

The X:A ratio does not have the final say in whether an animal develops as a male or a hermaphrodite. A number of genes necessary for interpreting the ratio have been identified through the isolation of mutations that cause worms to develop as members of the incorrect sex (Table 1). By analyzing double mutants it has been possible to determine how the various sex-determining genes interact. For example, null mutations in *fem-2* (feminization) cause XO and XX animals to develop as females, while *tra-1* (transformer) null mutations cause XO and XX animals to develop as males. Double mutants lacking both *fem-2* and *tra-1* activity develop as males (the *tra-1* mutation is epistatic to the *fem-2* mutation). Together these results suggest that *tra-1* has a feminizing role, and that *fem-2* acts to inhibit *tra-1* in XO but not XX animals. The complete set of epistatic interactions has led to the proposal that the sex-determining genes are part of a regulatory cascade, in which the genes at each level of the hierarchy negatively regulate those below them (Hodgkin 1980; Hodgkin 1987a). Subsequent molecular characterization of the genes has in many cases explained the observed genetic interactions, and a complex pathway combining the molecular and genetic information has been constructed (Figure 3). The components of this pathway and how they interact are discussed in more detail below.

1.6 The components that regulate sex determination and dosage compensation

1.6.1 Numerator elements in the X:A ratio

To date four regions of the X chromosome have been identified that behave as numerator elements (they act to increase the X:A ratio) (Akerib and Meyer 1994; Carmi et al. 1998). These have a feminizing role, since increasing their dose causes XO-specific feminization, while decreasing their dose causes XX-specific masculinization. Lethality is sometimes associated with changes in numerator dose, because the X:A ratio also regulates dosage compensation. Normally the level of transcription of the two hermaphrodite X chromosomes is down regulated so that the overall level of X-linked gene expression is similar to the male level (Meyer and Casson 1986; Donahue et al. 1987). If the X:A ratio is artificially skewed it can result in over or under expression of X-linked genes. A screen for suppressors of the male lethality caused by a duplicated numerator region led to the identification of the *fox-1* (feminizing on X) gene, which encodes an RNA-binding protein (Nicoll et al. 1997; Skipper et al. 1999). A second numerator gene, *sex-1* (signal element on X), was isolated in a search for mutations that cause the male-promoting sex determination gene *xol-1* (XO lethal) to be expressed in XX animals (Carmi et al. 1998). The predicted SEX-1 protein is similar to members of the nuclear hormone receptor family. Both FOX-1 and SEX-1 alter the expression of *xol-1* (described below). In the case of FOX-1 this regulation occurs posttranscriptionally (Nicoll et al. 1997; Skipper et al. 1999), while SEX-1 has been shown to bind the *xol-1* promoter, possibly acting as an inhibitor of transcription (Carmi et al. 1998). The genes responsible for the numerator activity of two other X chromosome regions remain to be identified, as do those that contribute to the denominator activity of the autosomes.

1.6.2 xol-1: the master switch gene

Animals homozygous for loss-of-function mutations in *xol-1* develop into hermaphrodites if they are XX, but die as embryos if they are XO (Miller et al. 1988). In normal males, *xol-1* is expressed at much higher levels than in hermaphrodites, a finding that is consistent with the sex-specific phenotypes (Rhind et al. 1995). Dead *xol-1* XO embryos often show signs of feminization, suggesting that they would have developed as hermaphrodites had the dosage compensation machinery not acted inappropriately. Indeed, the viability of XO mutant animals can be restored by a second mutation in the dosage compensation gene *dpy-27* (dumpy), and the resulting animals are often fertile hermaphrodites. Both the lethality and feminization of *xol-1* XO worms can be fully suppressed by mutations in the *sdc* (sex and dosage compensation) genes, suggesting that the role of *xol-1* is to inactivate the *sdcs* in the male. The sequence of the XOL-1 protein provides no clues as to how it acts, but there is evidence that it represses *sdc-2* directly by binding to its promoter (Rhind et al. 1995).

1.6.3 The sdcs: the branching point

Mutations in the *sdc* genes do not affect XO animals, but can cause XX lethality and masculinization (Villeneuve and Meyer 1990; Nusbaum and Meyer 1989; Delong et al. 1993). Higher levels of X-linked gene expression are detected in the mutant XX animals, consistent with the *sdcs* having a role in dosage compensation. *sdc-2* and *sdc-3* encode proteins that are localized to the hermaphrodite X chromosomes, as part of the dosage compensation complex (Dawes et al. 1999; Davis and Meyer 1997). The mechanism by which SDC-1 regulates X-linked gene expression is not clear, although its sequence suggests that it is a DNA binding protein (Nonet and Meyer

1991). The *sdc* genes likely promote the hermaphrodite sexual fate by repressing *her-1* (hermaphrodization), as transcript analysis shows that *her-1* is expressed at male levels in XX *sdc-1* and *sdc-2* animals (Trent et al. 1991), and SDC-2 associates with the *her-1* promoter (Dawes et al. 1999).

1.7 Sex determination in the soma

1.7.1 her-1 encodes a male-specific secreted protein

The *her-1* gene marks the beginning of the pathway that specifically regulates sex determination. It is dispensable for XX development, but XO animals lacking *her-1* activity are transformed into hermaphrodites (Hodgkin 1980). The HER-1 protein can act cell non-autonomously to promote the male fate (Hunter and Wood 1992), and it contains a secretory signal (Perry et al. 1993). These findings suggest that it acts as an extracellular ligand. Given the genetic evidence that *her-1* negatively regulates *tra-2* (Hodgkin 1980), and the fact that *tra-2* encodes a transmembrane protein (Okkema and Kimble 1991; Kuwabara et al. 1992), a reasonable model is that HER-1 promotes the male fate by binding directly to TRA-2. Consistent with this model is the observation that overexpression of TRA-2, or disruption of its putative HER-1 binding region, transforms XO animals into fertile hermaphrodites (Kuwabara and Kimble 1995; Hodgkin and Albertson 1995; Kuwabara 1996a).

1.7.2 tra-2 encodes a HER-1 receptor

The *tra-2* gene is required for hermaphrodite development—loss of function mutations transform XX animals into non-mating pseudomales, but do not affect XO animals (Klass et al. 1976; Hodgkin and Brenner 1977). In the absence of HER-1, TRA-2 is believed to promote the hermaphrodite fate by inhibiting the FEM proteins.

Overexpression of the cytoplasmic domain of TRA-2 feminizes XO animals (Kuwabara and Kimble 1995), and in the case of FEM-3 a direct interaction with this domain has been demonstrated (Mehra et al. 1999). *tra-3*, which acts at the same position as *tra-2* in the sex-determining gene hierarchy, encodes a protease that cleaves membrane bound TRA-2 (Sokol and Kuwabara 2000). The cleavage product that is released contains the FEM-3 binding site, thus the role of TRA-3 may be to generate a soluble feminizing fragment. The intracellular domain of TRA-2 has also been shown to bind TRA-1, the most downstream global regulator of somatic sex, and mutations in *tra-2* or *tra-1* that abolish this interaction cause weak somatic masculinization in XX animals (Lum et al. 2000). These observations have led to the proposal that these proteins act together as a complex to promote female development. Once again TRA-3 may be important, as the portion of TRA-2 it releases contains the TRA-1 binding site.

1.7.3 The FEM proteins bridge the gap between TRA-2 and TRA-1

Three *fem* genes are important for establishing the male fate, as null mutations in any one of them cause XO animals to develop as females (Doniach and Hodgkin 1984; Hodgkin 1986; Kimble et al. 1984). Based on genetic interactions, their most likely target in the soma is *tra-1* (Hodgkin 1987b). FEM-1 contains ankyrin repeats proposed to have a role in protein-protein interactions (Spence et al. 1990), FEM-2 has phosphatase activity (Chin-Sang and Spence 1996; Hansen and Pilgrim 1998), and FEM-3 is novel (Ahringer et al. 1992). The common position of the *fems* in the genetic pathway, and a demonstrated interaction in the case of FEM-3 and FEM-2 (Chin-Sang and Spence 1996), has led to the proposal that the three proteins function

as a complex. However, to date it is not known how the FEMs promote the male fate, either in terms of the way they interact, or the targets they act on.

1.7.4 tra-1 is the terminal regulator

TRA-1 is a zinc-finger protein that binds DNA *in vivo* (Zarkower and Hodgkin 1992; Zarkower and Hodgkin 1993). It is presumed to alter the expression of sex differentiation genes, which are distinguished from sex determination genes in that they do not specify the sexual fate of the entire soma or germ line. For example, *mab-3* (male abnormal) acts in XO animals to prevent expression of yolk protein genes and specify male sense organ differentiation (Shen and Hodgkin 1988). TRA-1 appears to reduce *mab-3* expression in hermaphrodites by binding to its promoter (Yi et al. 2000). Another known TRA-1 interaction involves the *egl-1* (egg laying) gene. TRA-1 represses *egl-1* transcription in two neurons required for hermaphrodite egg laying (Conradt and Horvitz 1999). In males active EGL-1 causes these neurons to undergo programmed cell death. Presumably other male-specific genes are repressed and hermaphrodite-specific genes activated by TRA-1, to account for the complete somatic sex reversal caused by TRA-1 mutations.

1.8 Sex determination in the germ line

1.8.1 Additional genes and different forms of regulation are involved

Some of the downstream components of the somatic pathway function differently in the germ line (Figure 4). For example, the *fem* genes sit at the bottom of the germ-line pathway, since unlike the situation in the soma, *tra-1* mutations do not restore the male fate in the absence of the *fems* (Hodgkin 1987b; Schedl et al. 1989). The germ line also employs additional genes, as mutations exist that cause inappropriate

feminization or masculinization of the germ line but not the soma. The regulatory mechanisms used to modulate gene activity also differ, in part because of the presence of the hermaphrodite sex, which requires the germ line to adopt one fate and then the other.

1.8.2 Establishing the male fate in the hermaphrodite

The sperm-producing phase in the XX germ line involves the inactivation of *tra-2*. Recall that in the soma the male fate is achieved through the binding of HER-1 to the extracellular domain of the TRA-2 protein. The mechanism of inactivation must be different in the XX germ line, because *her-1* transcripts cannot be detected (Trent et al. 1991), and loss-of-function *her-1* alleles do not affect hermaphrodite spermatogenesis (Hodgkin 1980). The characterization of gain-of-function mutations in *tra-2* that abolish sperm production in XX animals provided a clue as to how the gene is inhibited. The molecular lesions are located in the 3'UTR, in a region important for translational repression (Goodwin et al. 1993; Goodwin et al. 1997). One component of the repressor that binds this region may be encoded by *laf-1* (lethal and feminized). Animals with a reduced dosage of *laf-1* are feminized and overexpress transgenes containing the *tra-2* 3'UTR (Goodwin et al. 1997). Another regulator of *tra-2* translation is *gld-1* (defective in germ line development), a gene that is required for diverse aspects of germ-line development, including spermatogenesis in the hermaphrodite (Francis et al. 1995a, b). GLD-1 is similar to known RNA-binding proteins (Jones and Schedl 1995), it binds the *tra-2* 3'UTR *in vitro*, and TRA-2 is more abundant in the germ line of animals lacking *gld-1* activity (Jan et al. 1999). A third gene likely involved is *fog-2*. Like *gld-1*, it is required for hermaphrodite but not

male spermatogenesis (Schedl and Kimble 1988). Its protein product binds to GLD-1, forming a FOG-2/GLD-1/*tra-2* complex (Clifford et al. 2000). The reduction of *tra-2* activity may also occur post-translationally. A group of *tra-2* alleles that interfere with sperm production in the XX germ line contain missense mutations that disrupt a TRA-2/TRA-1 interaction. The increase in the feminizing ability of these alleles could be explained by the existence of a negative regulator that recognizes the TRA-2/TRA-1 complex, but not the unbound proteins (Lum et al. 2000).

Downstream of TRA-2 in the germ-line pathway are the *fems*, *fog-1*, *fog-3*, and *tra-1*. The primary target of TRA-2 appears to be FEM-3, since *fem-3(gf)* mutants undergo spermatogenesis throughout adulthood, and this phenotype is suppressed by *fog-2(lf)* mutations and by *tra-2(gf)* mutations (Barton et al. 1987). When *tra-2* is translationally repressed, *fem-3* and the other *fems* are able to promote spermatogenesis. Two germ line-specific genes, *fog-1* and *fog-3*, are also required, as loss-of-function mutations in either gene cause exclusive production of oocytes (Barton and Kimble 1990; Ellis and Kimble 1995). The FOG-1 protein shows similarity to CPEB (cytoplasmic polyadenylation element binding) proteins, which regulate the translation of specific mRNAs (Luitjens et al. 2000), while the sequence of FOG-3 provides little insight into how it might function (Chen et al. 2000). The targets of the FEMs, FOG-1, and FOG-3 are not known. The role of *tra-1* in the germ line is equally mysterious, and more complex. Gain-of-function mutations cause all germ cells to differentiate as oocytes, showing that constitutive *tra-1* activity can prevent spermatogenesis (Hodgkin 1987b). However, loss-of-function mutations reduce sperm production, suggesting that *tra-1* is needed to promote spermatogenesis

(Schedl et al. 1989). One explanation for this apparent discrepancy is that disruption of the interaction between TRA-2 and TRA-1 prevents proper regulation of TRA-2 (Wang and Kimble 2001).

1.8.3 Switching to the female fate in the hermaphrodite

After about 300 sperm are produced in the hermaphrodite germ line there is a switch to the female fate. Mutations in the 3'UTR of *fem-3* can cause constitutive sperm production (Barton et al. 1987; Ahringer and Kimble 1991), suggesting that the switch may be achieved by translational repression. One potential repressor was identified using the three-hybrid system. It is referred to as FBF (*fem-3* binding factor), and RNAi indicates that hermaphrodites are unable to make the sperm-oocyte switch in its absence (Zhang et al. 1997). FBF functions as a complex with NOS-3, a protein that resembles *Drosophila* Nanos (Kraemer et al. 1999). Two additional Nanos-related proteins (NOS-1 and NOS-2) participate in the switch but do not interact with FBF directly (Kraemer et al. 1999). Mutations in six other genes, *mog-1* to *mog-6*, yield phenotypes similar to that of the *fem-3(gf)* alleles (Graham and Kimble 1993; Graham et al. 1993), and enhance the expression of transgenes regulated by the *fem-3* 3'UTR (Gallegos et al. 1998). MOG-1, MOG-4, and MOG-5 are similar to the DEAH-box RNA helicases that regulate RNA splicing in yeast (Puoti and Kimble 1999; Puoti and Kimble 2000). It is not known whether the MOGs bind to *fem-3* mRNA directly, or if they alter the splicing of other genes involved in *fem-3* repression.

1.8.4 Germ-line sex determination in the male

The XO germ line adopts the male fate using a different mechanism than XX animals. *her-1* is dispensable for sperm production in XX animals, but if males carrying a

temperature sensitive allele of *her-1* are shifted to the restrictive temperature, oocytes are produced instead of sperm (Schedin et al. 1994). The germ-line role of HER-1 is thought to be the same as its role in the soma—to bind and inactivate TRA-2 (Kuwabara and Kimble 1992). *fog-2* and *gld-1*, which are required for translational repression of *tra-2* in XX animals, are not needed for male spermatogenesis (Schedl and Kimble 1988; Francis et al. 1995a). Thus the interaction between HER-1 and TRA-2 may be the primary form of *tra-2* inactivation. The remaining components of the pathway are thought to act as they do during the XX stage of spermatogenesis.

1.9 The evolution of sex determination

Interspecies comparisons are now being used to study the evolution of sex-determining genes and pathways, and thus far some surprising results have been obtained.

1.9.1 Sex-determining genes evolve rapidly at the sequence level

Comparisons of sex-determining genes between closely related species suggest that they evolve more rapidly than other genes. The orthologs of four *C. elegans* sex-determining genes have been isolated from the hermaphroditic species *C. briggsae* (de Bono and Hodgkin 1996; Kuwabara 1996b; Hansen and Pilgrim 1998; Streit et al. 1999). In each case the predicted proteins have diverged more than the average level of divergence observed for 11 non-sex-determining proteins isolated from the same species (de Bono and Hodgkin 1996). Similar findings have been obtained from comparisons between mammalian species, and between *Drosophila* species (Whitfield et al. 1993; Tucker and Lundrigan 1993; O'Neil and Belote 1992; McAllister and McVean 2000). Often the poorly conserved regions of a protein are thought to lack

functional importance, but in the case of sex-determining proteins it has been proposed that positive selection could be causing functionally important regions to diverge rapidly (Whitfield et al. 1993; Walthour and Schaeffer 1994). Positive selection has been demonstrated in the case of proteins that mediate sperm to egg binding in marine environments, where rapid divergence could facilitate reproductive isolation (Metz and Palumbi 1996; Swanson and Vacquier 1998). In the absence of positive selection proteins are said to evolve neutrally, with divergence occurring because it does not interfere with function. A recent study of the *transformer* sex-determining gene in *Drosophila* suggests that although it is evolving at a rapid pace, the changes are neutral in function (McAllister and McVean 2000). Why would one neutrally evolving protein diverge faster than another? One explanation is that the more quickly diverging protein contains more residues that are not important for function. Another is that the activity of the quickly diverging protein can vary more without having a detrimental effect on the fitness of the organism.

1.9.2 The C. elegans *and* Drosophila *pathways are not related*

A recurring theme in biology is that different organisms use similar genetic components to regulate development. Many scientists are quick to point out this genetic conservation when justifying their work in model organisms to non-scientists and granting agencies. In the case of sex determination, information gained in one species may not always be applicable to another. Comparisons of the proteins comprising the *C. elegans* and *Drosophila* pathways have revealed almost no significant similarity, with each species relying on distinct biochemical mechanisms to propagate the sexual fate decision (for review see Cline and Meyer 1996). The

somatic pathway in *C. elegans* involves negative regulatory interactions at the protein level, while the *Drosophila* somatic pathway involves a series of positive interactions at the level of RNA splicing (Figure 5). In the worm the somatic sex-determining genes all participate in the germ-line fate decision, while in the fly most somatic sex-determining genes have no germ-line role. To date the only molecular similarity involves the *C. elegans mab-3* gene, and the terminal sex-determining gene in *Drosophila*, *doublesex* (Raymond et al. 1998). From the limited information available, the mouse pathway lacks similarity to the worm or fly pathways, apart from a homolog of *mab-3*/*doublesex* that is required for testis differentiation (Raymond et al. 2000). It has been proposed that the most downstream components of the pathways are the most ancient, and that the upstream components have been assembled independently (Wilkins 1995).

1.10 Using FEM-2 to understand sex-determining gene and pathway evolution

In this study the evolution of the *C. elegans* protein FEM-2 is examined, in order to gain further insights into how sex-determining proteins and pathways evolve. FEM-2 is particularly suited for this analysis for a variety of reasons. It is a member of the protein phosphatase type 2C (PP2C) family of serine/threonine phosphatases (Pilgrim et al. 1995; Chin-Sang and Spence 1996; Hansen and Pilgrim 1998), of which there are other members in *C. elegans* with no known role in sex determination. These FEM-2 paralogs can be used as internal controls to determine the rate and pattern of sequence divergence that is typical for PP2C proteins. Also, the three-dimensional structure of a human PP2C is known (Das et al. 1996), allowing sequence conservation and structure/function information to be integrated. The study of gene

and gene pathway evolution in *C. elegans* has practical advantages, as homologs in related species can be silenced using RNA interference, and transgenes from one species can be tested for function in another.

1.10.1 A brief history of fem-2

The *fem-2* gene has a well-established role in *C. elegans* sex determination. It was first described as the temperature-sensitive allele *b245* (Kimble et al. 1984). Homozygous *b245* worms develop normally if they are raised at 16°C, but are feminized if they are raised at 25°C. In XX animals the feminization is complete, with the germ line producing only oocytes. In XO animals both sperm and oocytes are produced, and the somatic gonad is feminized. The mutant phenotypes are observed among the progeny of homozygous *fem-2* parents. However, homozygous *fem-2* worms arising from a *fem-2*/+ mother do not show the complete mutant phenotype, indicating that *fem-2* activity is supplied maternally.

A stronger allele of *fem-2*, *e2105*, was isolated by Hodgkin (1986) as a suppressor of the masculinization caused by a *tra-3* mutation. Homozygous *e2105* XX animals are unable to produce sperm at all temperatures, while the XO phenotype is temperature-sensitive. Mutant XO animals raised at 25°C develop into fertile females, while those raised at 20°C develop into intersex animals. The intersexes have an incomplete male tail, a two-armed gonad, and a partially formed vulva. Their germ line is also feminized, producing oocytes instead of sperm. The mutant phenotypes are not observed in homozygous progeny from heterozygous mothers, again indicating that *fem-2* mRNA or protein is transferred from mother to offspring.

The position of *fem-2* on the genetic and physical maps of the *C. elegans* genome was refined (Pilgrim 1993), and the gene was precisely identified using a complementation assay, in which DNA in the vicinity of the gene was injected into *b245* animals (Pilgrim et al. 1995). Sequence comparisons suggested that *fem-2* encodes a PP2C, and *in vitro* tests confirmed that FEM-2 is a member of this protein family (Chin-Sang and Spence 1996; Hansen and Pilgrim 1998). With the wild-type sequence available, the mutant alleles of *fem-2* could be characterized (Figure 6). The *b245* allele contains a single missense mutation, which alters a residue conserved among a wide variety of PP2Cs and results in a protein with greatly reduced phosphatase activity (Hansen and Pilgrim 1998). The *q117* allele, which yields mutant phenotypes that resemble those produced by *b245*, contains a missense mutation that alters a residue in the amino-terminal domain of the protein, upstream of the PP2C domain. It is not known if the resulting protein possesses phosphatase activity. Finally, the *e2105* allele is altered such that it contains a premature stop codon, which should result in a protein that lacks most of the PP2C domain.

1.10.2 The roles of other PP2Cs

Serine /threonine phosphatases (PPases) have been classified into four families based on their biochemical properties. PP1 and PP2A do not require metal ions and are sensitive to okadaic acid and microcystins; PP2B requires Ca^{2+} for full activity; and PP2C requires Mg^{2+} or Mn^{2+} (Cohen 1989). Subsequent sequence comparisons placed these groups into two distinct gene families, designated PPP and PPM (Barford 1996). PPP consists of PP1, PP2A, and PP2B, and PPM consists of the PP2C enzymes. In mammals the PP2C family includes PP2Cα (Tamura et al. 1989; Mann et al. 1992),

PP2Cβ (Wenk et al. 1992), PP2Cγ/FIN13 (Guthridge et al. 1997; Travis and Welsh 1997; Murray et al. 1999), PP2Cδ/Wip1 (Tong et al. 1998; Fiscella et al. 1997), ILKAP (integrin-linked kinase-associated phosphatase) (Leung-Hagesteijn et al. 2001), and CaMKPase (Ca^{2+}/calmodulin-dependent kinase II phosphatase) (Kitani et al. 1999; Tan et al. 2001). In the case of human PP2Cα the three-dimensional structure has been determined (Das et al. 1996), and the residues that are likely critical for catalytic activity have been identified (Figure 7). In *C. elegans* the PP2C family consists of three predicted proteins (T23F11.1, F25D1.1, and F42G9.1) and FEM-2.

PP2C substrates have been identified through the purification of cell fractions that dephosphorylate a protein of interest (Cheng et al. 2000; Hishiya et al. 1999), by immunoprecipitation (Zhu et al. 1999), and by performing a two-hybrid screen (Strovel et al. 2000). Most of the known interactions involve human PP2Cα, which appears to act on several different substrates (Table 2). In genetically-tractable organisms, lesions in PP2C genes have been associated with mutant phenotypes (Table 3). In most cases the substrates regulated by the products of these genes are not known. From the molecular and genetic information obtained thus far it is clear that PP2Cs regulate a variety of biological pathways. Currently, FEM-2 is the only PP2C known to regulate sex-determination.

1.11 The main objectives of this work

The primary goal of this work was to isolate and characterize the *fem-2* gene from the male/female nematode species CB5161, so that interspecies comparisons of sequence and function could be made. Such comparisons have the potential to shed light on how sex-determining pathways, and regulatory pathways in general, diverge between

species. Knowledge of pathway evolution is important, because it could influence how we apply biological information gained in model systems to other organisms including ourselves. Specific objectives of this study were to:

1. Isolate the orthologs of the *C. elegans* PP2Cs *fem-2*, T23F11.1, and F25D1.1 from CB5161.
2. Determine whether *fem-2* has evolved more quickly than T23F11.1 and F25D1.1.
3. Isolate PP2C sequences from zebrafish and humans to determine which regions of the PP2C domain typically diverge most rapidly.
4. Test *fem-2* transgenes from *C. briggsae* and CB5161 for function in *C. elegans*, to determine whether functionally important portions of the proteins have diverged.
5. Use RNAi to determine whether *fem-2* is a component of the sex-determining pathways in *C. briggsae* and CB5161.
6. Identify potential FEM-2 substrates in *C. elegans* using the two-hybrid system.
7. Determine the expression pattern of FEM-2 using reporter constructs and anti-FEM-2 antisera.
8. Investigate the function of a human homolog of *fem-2*.

2. Materials and Methods

2.1 Nematode strains

The following wild-type nematode strains were used in this study: *C. elegans* strain N2, *C. briggsae* strain AF16, and *Caenorhabditis* sp. strain CB5161. The Caenorhabditis Genetics Center and others originally referred to CB5161 as *C. remanei*, but it is now known to represent a distinct species (Thomas and Wilson 1991; Fitch et al. 1995). It is not known whether CB5161 is more closely related to *C. elegans* or *C. briggsae*.

2.2 PCR conditions and primers

The sequence of each primer referred to in the Materials and Methods section is given in Table 4. Additional primers used in this study and by others in the Pilgrim lab appear in Figures 8, 9, and 10, which show the annealing sites of the primers in relation to the DNA sequence and protein translation of their templates. PCR was performed in a custom buffer (60 mM Tris-SO_4, pH 9.1; 18 mM $(NH_4)_2SO_4$; 2.6 mM $MgSO_4$) using the following program: 94°C for 2 minutes, 55°C for 2 minutes, 72°C for 2 minutes, one cycle; 94°C for 1 minute, 55°C for 2 minutes, 72°C for 2 minutes, 35 cycles. In some cases the length of the extension time was modified to promote the production of shorter or longer products. For the BJM1 and BJM2 primers the annealing temperature was lowered to 50°C.

2.3 Isolation of PP2C homologs by PCR

Degenerate oligonucleotide primers (BJM1 and BJM2) were designed based on conserved PP2C sequences detected through the alignment of FEM-2 with other known and predicted PP2Cs. The annealing sites of these primers correspond to

portions of PP2C motifs two and eight, of the 11 conserved motifs detected in a detailed sequence analysis of the PP2C superfamily (Bork et al. 1996). To isolate PP2Cs from CB5161, PCR was performed using genomic DNA as the template. Products were separated on an agarose gel, and a band corresponding to approximately 650 bp was excised from the gel and the DNA was isolated using the Sephaglas BandPrep Kit (Pharmacia Biotech). The DNA was then cloned into pGEM-T (Promega) and six clones were examined by DNA sequencing. Two of the clones, which appeared to be identical, were found to be similar to *C. elegans fem-2* (pDP#PSr1 and pDP#PSr3), while another was similar to the predicted *C. elegans* PP2C gene T23F11.1 (pDP#PSr2). The remaining three sequences did not appear to be PP2C related and were ignored. The PCR and cloning procedures were repeated using the primers F25FI and F25RO, which are designed to anneal to the predicted *C. elegans* PP2C gene F25D1.1. Two clones were sequenced and one (pDP#PS3-1p.755) was found to contain the putative CB5161 F25D1.1 ortholog. To obtain PP2C sequences from zebrafish, total RNA was isolated from adult animals using RNAzol B as directed by the supplier (Tel-Test), and it was reverse transcribed by SuperScript II reverse transcriptase (Life Technologies) using oligo-dT as a primer. The resulting cDNA was amplified using the BJM1 and BJM2 primers, and the products were separated on an agarose gel. A band corresponding to approximately 500 bp in size was excised, cloned, and characterized as described above. Three different PP2C sequences were obtained, and based on their similarity to mammalian PP2Cs they were named *pp2cα1*, *pp2cα2*, and *pp2cβ*.

2.4 Isolation of the complete CB5161 *fem-2* cDNA

Primers for RT-PCR were designed based on the partial genomic sequence of *Cs-fem-2* isolated as described above. To obtain the *Cs-fem-2* cDNA, total RNA was isolated from adult CB5161 worms as described (Hansen and Pilgrim 1998) and reverse transcribed by SuperScript II reverse transcriptase (Life Technologies) using the AP primer (Life Technologies). The resulting single stranded cDNA served as the template in subsequent PCR reactions. A DNA copy of the 3' end of the *Cs-fem-2* transcript was obtained by performing PCR using the UAP and CspefemFU primers. A nested primer (CspefemFD) was then used with UAP for further amplification, and the resulting product was cloned into pGEM-T (Promega) and sequenced. To obtain the 5' end of the transcript, PCR was performed using the primer CspefemRD in conjunction with the LVSL1 primer, which is complementary to the spliced leader (SL1) that is present on the 5' end of many *C. elegans* transcripts (Blumenthal and Steward 1997). The nested primer CspefemRU was then used with LVSL1 for further amplification, and the resulting product was cloned into pGEM-T and sequenced. A DNA copy of the middle portion of the *Cs-fem-2* transcript was obtained by performing RT-PCR with the degenerate primers BJM1 and BJM2. A product of approximately 500 bp was isolated from an agarose gel and cloned into pGEM-T, and four clones were sequenced. Two were found to correspond to *Cs-fem-2*, while the others matched Cs-T23F11.1. The cDNA sequences obtained using PCR overlap to give a complete cDNA that is identical to the genomic sequence (described later), apart from absent intronic sequences, a poly-A tail, an inferred SL1 sequence, and six single base differences between the genomic and cDNA sequence. These single base changes were likely introduced during PCR amplification or reverse transcription of

the cDNA, and hence the genomic sequence is considered to be the correct sequence. These differences between the cDNA and genomic DNA do not affect the protein sequence of the PP2C domain segment used to assess the rate of FEM-2's evolution relative to other PP2Cs. Although no signs of alternative splicing were observed during the amplification and cloning of the cDNA fragments, the possibility that *Cs-fem-2* is alternatively spliced cannot be excluded.

2.5 Northern analysis of *Cs-fem-2* expression

Northern analysis of *Cs-fem-2* expression was performed with the assistance of Dave Hansen (University of Alberta). Gel electrophoresis of RNA and northern blotting were performed as described (Hansen and Pilgrim 1998). A 1.1% agarose gel was used to electrophorese five micrograms of total RNA, which was then transferred to GeneScreen Plus nylon membrane (New England Biolabs). A ^{32}P-labeled DNA probe prepared from the insert of clone pDP#PSr1 was hybridized to the blot using Hybrisol II hybridization solution (Oncor) overnight at 65°C. The blot was washed twice with 2X SSC for 15 minutes each, once with 2X SSC, 0.1% SDS for 30 minutes, and once with 0.1X SSC, 0.1% SDS for 10 minutes.

2.6 Obtaining a genomic clone of *Cs-fem-2*

A genomic clone of *Cs-fem-2* was obtained with the assistance of Dave Hansen (University of Alberta). Southern blotting of genomic CB5161 DNA revealed a 6-kbp *Sst*I band that hybridizes to the insert from pDP#PSr1. A mini-library in pBluescript II SK- (Stratagene) was constructed from gel purified *Sst*I-digested genomic DNA 5 to 7 kbp in size, and the library was screened with the same probe. Partial sequence of a positive clone, pDP#DH150, revealed that it did not contain the entire coding region.

A partially overlapping clone, pDP#DH160, was isolated from a similar library prepared from a size selected *Bam*HI/*Hin*dIII digest. A clone containing the entire *Cs-fem-2* coding sequence was constructed by combining portions of pDP#DH150 and pDP#DH160. Correct orientation of the insert in the resulting clone, pDP#DH161, was confirmed by sequencing.

2.7 Construction of chimeric *fem-2* genes

Unique *Xba*I and *Xho*I sites were added to a wild-type genomic *Ce-fem-2* clone (pDP#DH11; Hansen and Pilgrim 1998) using *in vitro* mutagenesis (Leatherbarrow and Fersht 1986) and the primers Cele3Xba and Cele5Xho. The resulting plasmid (pDP#DH91) contains an *Xho*I cleavage site 11 bases upstream of the *Ce-fem-2* start codon, and an *Xba*I cleavage site 6 bases upstream of the stop codon. To prepare a Ce/Cs *fem-2* fusion, PCR was performed using pDP#DH161 as the template and the primers Cspe5Xho and Cspe3Nhe, which introduce an *Xho*I cleavage site 5 bases upstream of the *Cs-fem-2* start codon, and an *Nhe*I cleavage site 1 base downstream of the stop codon. The Cspe3Nhe primer also introduces a silent substitution to remove an existing *Xho*I site. The PCR product was digested with *Xho*I and *Nhe*I and ligated to the promoter-containing fragment of pDP#DH91, which was generated by an *Xho*I/*Xba*I digest. The resulting clone (pDP#PS3-1p.805) was sequenced to confirm that the expected changes were present. To generate a Ce/Cb *fem-2* fusion, PCR was performed using pDP#DH53 (Hansen and Pilgrim 1998) as the template and the primers Cbrig5Xho and Cbrig3Xba. These introduce an *Xho*I cleavage site 5 bases upstream of the *Cb-fem-2* start codon, and an *Xba*I cleavage site 14 bases downstream of the stop codon. The PCR product was digested with *Xho*I and *Xba*I and ligated to

the promoter-containing fragment of pDP#DH91, which was generated using the same enzymes. The resulting clone (pDP#PS5-1p.805) was sequenced to confirm that the expected changes were present.

2.8 Transformation of *fem-2* animals

The following plasmids were assayed for their ability to rescue *fem-2* mutants: pDP#DH11 (*Ce-fem-2*; Hansen and Pilgrim 1998), pDP#DH161 (*Cs-fem-2*), pDP#DH53 (*Cb-fem-2*; Hansen and Pilgrim 1998), pDP#PS3-1p.805 (*Cs-fem-2* with *Ce-fem-2* regulatory regions), and pDP#PS5-1p.805 (*Cb-fem-2* with *Ce-fem-2* regulatory regions). To test for somatic rescue, plasmids were injected with pRF4 (*rol-6*(*su1006dm*)) into the gonads of *C. elegans* hermaphrodites of strain DP51 (*fem-2*(*e2105*) *unc-45*(*r450ts*)/*sC1* [*dpy-1* (*s2171*)]; *him-8*(*e1489*)) as described (Mello et al. 1991). Rolling Unc hermaphrodites were transferred to separate plates and allowed to self-fertilize at 20°C. Unc m-z- (i.e. lacking *Ce-fem-2* product from both mother and zygote) hermaphrodite (XX) and male (XO) rolling animals were raised at 20°C or 25°C and compared to control DP51 animals. To test for germ-line rescue, homozygous *fem-2(b245)* animals, which do not produce sperm when raised at 25°C, were injected with the same plasmids. Progeny from the injected animals were raised at 25°C, and rolling animals were examined for the presence of eggs.

2.9 RNA-mediated interference

Complete or near-complete *fem-2* cDNAs from *C. elegans*, *C. briggsae*, and CB5161 were cloned into pBluescript (Stratagene). PCR using the primers T7big and T3becomesT7 was performed on each construct to incorporate T7 promoter sequences into both strands of the cDNA. RNA was synthesized from the PCR products using

the MEGAscript *in vitro* transcription kit (Ambion), and injected at a concentration of 1 mg/ml into the gonad or gut of *C. elegans*, *C. briggsae*, or CB5161 worms. Injected worms were transferred to separate plates after approximately 20 hours, and their progeny were examined once they reached the adult stage. Injected CB5161 worms were always kept with five males. All worms were raised at 25°C.

2.10 Searching for polymorphisms in *fem-2* and F25D1.1

Eight additional geographic isolates of *C. elegans* were obtained from the Caenorhabditis Genetics Center (Table 5). Genomic DNA corresponding to the *fem-2* gene was amplified from each strain by PCR using the primers AMC3 and SAD2 as described (Pilgrim et al. 1995). A region of the F25D1.1 gene was amplified in a separate set of reactions using the primers F25FI and F25RI. The PCR products were separated by gel electrophoresis and purified using the Sephaglas BandPrep Kit (Pharmacia Biotech). The purified *fem-2* and F25D1.1 DNA was sequenced using the nested primers DHA19 and F25D1seq, respectively.

2.11 Two-hybrid screen for proteins that interact with FEM-2

A *Bgl*II fragment encoding the amino-terminal domain of FEM-2 was obtained from the plasmid pDP#DH022 (Dave Hansen, University of Alberta). The fragment was ligated to the yeast two-hybrid bait plasmid pAS1-CYH2 (Stephen Elledge, Baylor College of Medicine), which had been digested with *Bam*HI. A clone containing the insert in the desired orientation (pDP#PS22) was transformed into the yeast strain PJ69-4A (Philip James, University of Wisconsin). The *C. elegans* cDNA library λACT-RB1 (Robert Barstead, Oklahoma Medical Research Foundation) was transformed into the bait-carrying strain, and cells were grown on incomplete media to

select for the plasmids and reporter gene expression. Colonies that appeared within five days were streaked on complete media, to allow for plasmid loss. Prey plasmids were then isolated from colonies that had lost the bait but not the prey plasmid, as indicated by growth on leucine-deficient but not tryptophan-deficient media. The positive clones were retransformed into strains carrying the original bait and pSE1112, which encodes a GAL4-SNF1 fusion protein. Plasmids that supported colony growth with the *fem-2* bait but not the SNF1 bait were sequenced.

2.12 Construction and expression of *fem-2* reporter genes

A unique *Bgl*II site was added to a wild-type genomic *Ce-fem-2* clone (pDP#DH11; Hansen and Pilgrim 1998) using *in vitro* mutagenesis and the primer fem-2-Bgl-fast (Leatherbarrow and Fersht 1986). Presence of the expected single base change in the resulting plasmid (pDP#PS1p.644) was confirmed by restriction digestion and sequencing. *Bgl*II-digested pDP#PS1p.644 was then used in three separate ligation reactions to make three reporter constructs. A 128 bp *Bgl*II fragment from the plasmid pUR#91 (Morris Maduro, University of Alberta) was used to make a hemagglutinin (HA) fusion (pDP#PS4-2p.648), a 600 bp *Bgl*II fragment from the plasmid pDP#DH76 (Dave Hansen, University of Alberta) was used to make a GFP fusion (pDP#PS2-3p.648), and a 3700 bp fragment from the plasmid pDP#DH71 (Dave Hansen, University of Alberta) was used to make a *lacZ* fusion (pDP#PS3-2p.648). In each case plasmids containing a single insert in the desired orientation were identified by restriction analysis followed by sequencing. The *fem-2* fusions and pDP#PS1p.644 were injected separately with pRF4 (*rol-6*(*su1006dm*)) into the gonads of wild-type and *fem-2(b245)* worms as described (Mello et al. 1991). The ability of the constructs

to supply *fem-2* activity was assayed by examining rolling *fem-2(b245)* animals raised at 25°C for the presence of eggs. Expression of the GFP and β-GAL reporters was analyzed in whole animals using fluorescence microscopy and X-gal staining, respectively. To detect HA expression, protein was extracted from worms and separated by SDS-PAGE. Western blotting and detection were then performed using the protocols and reagents included in the ECL kit (Amersham). FEM-2 was detected using rabbit anti-FEM-2 polyclonal antibodies (Petra Jackle, University of Alberta), followed by anti-rabbit IgG conjugated to horseradish peroxidase (Amersham). HA-tagged protein was detected using mouse anti-HA monoclonal antibody (Roche Molecular Biochemicals), followed by anti-mouse IgG conjugated to horseradish peroxidase (Amersham).

2.13 Characterization of a human homolog of *fem-2*

A human cDNA (KIAA0015) encoding a predicted PP2C was obtained from Nobu Nomura (Kazusa DNA Research Institute). The gene corresponding to the cDNA had been localized to chromosome 22 using human-rodent hybrid cell lines and PCR (Nomura et al. 1996). To obtain a copy of the gene and to refine its position on the physical map, a chromosome 22 cosmid library (LL22NC03; Lawrence Livermore National Laboratory) was screened by hybridization. The probe was generated using the primers DHA24 and DHA25, which amplify a coding portion of the cDNA. Expression of KIAA0015 transcripts was analyzed (with the assistance of Graham Banting, University of Alberta) by hybridization using a multiple tissue northern blot (Clontech). The probe was generated by digestion of the KIAA0015-containing

plasmid with *Hin*dIII and *Not*I. The resulting fragment included the entire 5134 bp cDNA.

2.14 Characterization of a zebrafish pyruvate dehydrogenase phosphatase

When the BJM1 and BJM2 primers were used to amplify zebrafish PP2C sequences, a 650 bp product was obtained in addition to the expected 500 bp product. A clone derived from the larger product (pDP#PSB5p.399) was sequenced, and one of its predicted translations was found to resemble rat and bovine pyruvate dehydrogenase phosphatase (PDP) proteins. Of the two known mammalian isoforms, PDP2 was the most similar, thus the zebrafish gene was named *pdp2*. Using the putative *pdp2* cDNA as a probe, Angela Manning (University of Alberta) screened a 15-19 hpf (hours post fertilization) zebrafish cDNA library (Bruce Appel, University of Oregon). A positive clone (pDP#PSXhoI#1p.549) was sequenced and found to contain a complete or near complete *pdp2* cDNA. The *pdp2* gene was mapped using the radiation hybrid panel LN54 (Neil Hukriede, National Institutes of Health) and the primers PDPfor and PDPrev. Each of the 93 cell lines comprising the panel was tested in two separate PCR reactions, as recommended by the suppliers. *pdp2* expression was analyzed by Angela Manning (University of Alberta), using northern blots prepared from adult zebrafish and oocyte RNA. The probe used in the hybridizations was obtained from pDP#PSB5p.399 as a 600 bp *Pst*I/*Sst*II fragment.

2.15 Sequence analysis and manipulation

Sequences were aligned using CLUSTAL X with the default settings (Thompson et al. 1997). The multiple alignment showing the complete FEM-2 sequences with Ce-T23F11.1 and Ce-F25D1.1 was constructed by aligning the PP2C and the flanking

regions separately. To estimate the number of synonymous (K_s) and nonsynonymous (K_a) substitutions per nucleotide site, orthologous PP2C domain protein sequences were aligned, and then the corresponding nucleotide sequence alignments were prepared manually. These were analyzed using K-Estimator 5.5 (Comeron 1999), which estimates K_s and K_a using the method described by Comeron (1995). The correction for multiple substitutions described by Kimura (1980) was used when applicable, otherwise the Jukes and Cantor (1969) correction was applied. A phylogenetic tree was generated from aligned PP2C domains using CLUSTAL X and the neighbor-joining method (Saitou and Nei 1987). Portions of the alignment containing gaps were excluded from the analysis, and the correction for multiple substitutions was performed. The results were analyzed using the bootstrap method (1000 replicates) to provide confidence levels for the tree topology. To color the three-dimensional structure of PP2Cα according to the sequence conservation of FEM-2, the Ce-FEM-2, Cb-FEM-2, and Cs-FEM-2 PP2C domain sequences were aligned with the PP2Cα sequence. Each residue in the solved structure was then assigned a color according to the level of conservation among the aligned residues in the FEM-2 proteins, using COMBOSA3D (Stothard 2001). The PP2Cα alignment procedure was repeated using the PP2C domains from Ce-FEM-2, Ce-F25D1.1, and Ce-T23F11.1, so that the structure could be colored to show sequence conservation among FEM-2 and its paralogs. The coloring procedures were repeated using the predicted structure of the Ce-FEM-2 PP2C domain, as modeled by SWISS-MODEL (Guex and Peitsch 1997). When using the predicted structure the comparison sequences were aligned with the PP2C domain of Ce-FEM-2 instead of PP2Cα. The

same localization of conserved residues was obtained, and only the comparisons using the solved PP2Cα structure are presented. Sequence alignment figures were colored using the Sequence Manipulation Suite (Stothard 2000), and the figures showing primer annealing sites were generated using the Sequence Extractor (Stothard unpublished; http://www.bioinformatics.org/seqext/).

3. Results

Most of the results described in this thesis have been published in a similar form elsewhere (Stothard et al. 2002).

3.1 Isolation of *Cs-fem-2*

All previous studies of sex-determining gene evolution in *Caenorhabditis* have relied on low stringency hybridization or cloning by conserved synteny to isolate sex-determining gene orthologs (de Bono and Hodgkin 1996; Kuwabara 1996b; Hansen and Pilgrim 1998; Streit et al. 1999; Haag and Kimble 2000). In this study degenerate PCR primers were designed to anneal to two conserved PP2C motifs, and these were used to isolate a genomic fragment of the CB5161 homologue of *fem-2*. Additional primers were designed based on the sequence of the genomic fragment, and these were used to obtain the complete *Cs-fem-2* cDNA using 3' RACE and RT-PCR. Cloning of the 5' end of the transcript was facilitated by an SL1 spliced leader sequence, which is also present on *Ce-fem-2* and *Cb-fem-2* transcripts (Pilgrim et al. 1995; Hansen and Pilgrim 1998). The size of the complete cDNA is consistent with the size detected by northern blotting (Figure 11). Single product bands were obtained using RT-PCR, and northern analysis shows a single band of hybridization. Thus there is no evidence for alternative splicing of *Cs-fem-2*, which is also the case for *Ce-fem-2* and *Cb-fem-2* (Pilgrim et al. 1995; Hansen and Pilgrim 1998). A complete genomic copy of *Cs-fem-2* was isolated from a CB5161 minilibrary probed with the sequence fragment obtained using PCR (Figure 12A). Comparison of the genomic sequence to the cDNA sequence reveals the presence of seven introns, nearly twice as many as *Ce-fem-2* (Figure 12B). Three of the introns are located in the same positions in all three *fem-2*

genes, and only one of the introns in *Cs-fem-2* is not found in either *Cb-fem-2* or *Ce-fem-2* (Figure 13). In contrast, *Cs-fem-2* does not share any intron positions with the *C. elegans* PP2C sequences T23F11.1 or F25D1.1 (Figure 13).

3.2 Comparison of Cs-FEM-2 to Ce-FEM-2 and Cb-FEM-2

The protein predicted from the *Cs-fem-2* cDNA and genomic sequences consists of 483 amino acids, 34 amino acids longer than Ce-FEM-2 and 19 amino acids shorter than Cb-FEM-2. Cs-FEM-2 is 59% identical to Ce-FEM-2, which is similar to the conservation seen when Ce-FEM-2 is compared with Cb-FEM-2, and when Cs-FEM-2 is compared to Cb-FEM-2. The residues that are mutated in two temperature-sensitive alleles of *fem-2* in *C. elegans* are conserved, as is a residue associated with functional mutations in two *Arabidopsis* PP2Cs (Figure 13). Each protein also contains the seven metal binding and phosphate binding residues predicted by analysis of the three-dimensional structure of human PP2Cα (Das et al. 1996). In all three species FEM-2 contains a long amino-terminal domain that is poorly conserved and absent from many other PP2Cs. There is also a group of acidic amino acids near the carboxy terminus that is not a general characteristic of PP2Cs (Bork et al. 1996).

3.3 FEM-2's PP2C domain evolves faster than the same domain in other PP2Cs

To compare the evolution of FEM-2 to the evolution of other PP2Cs, additional PP2C sequences were isolated from CB5161 and the zebrafish using PCR. The sequences were then aligned with their best matches in the GenBank database, as detected by performing BLAST searches (Altschul et al. 1997) (Figure 14). The PP2C domain of FEM-2 was found to be the most diverged of the PP2Cs obtained, showing 68% identity between *C. elegans* and CB5161 (Table 6). The same domains in the *C.*

elegans genes F25D1.1 and T23F11.1 were found to be 90% and 95% identical to their CB5161 orthologs, respectively. To better compare the rates of evolution of the PP2C domains, and to summarize the ortholog/paralog relationships, a phylogenetic tree was constructed (Figure 15). The branch lengths of the tree, which serve as a better indication of evolutionary distance than simple percent differences in amino acid sequence, also indicate that the FEM-2 orthologs are more diverged than the other PP2C orthologs. The bootstrap values for the branches show that all three of the PP2C sequences from CB5161 can be unambiguously paired with their *C. elegans* orthologs. Thus the phylogenetic analysis of the predicted protein sequences, along with the conserved intron/exon boundaries, indicate that *Cs-fem-2* is the probable sequence ortholog of *Ce-fem-2*. From the limited sequence data available, the precise relationship between the species CB5161, *C. elegans*, and *C. briggsae* cannot be determined.

3.4 FEM-2 has a unique catalytic region that is conserved

When each non-FEM-2 partial PP2C domain sequence is aligned with its ortholog one of the most diverged regions is located between motifs three and four, while the most conserved region spans motif six (Figure 14). Although the motif six region is perfectly conserved within ortholog groups, between ortholog groups it differs at many sites, suggesting that it may convey specificity to the PP2C domain. The pattern of substitutions between Cs-FEM-2 and Ce-FEM-2 is similar to the pattern observed for the non-sex-determining PP2Cs. Within the FEM-2 group the region between motifs three and four has diverged, while motif six is completely conserved. The crystal structure of human PP2Cα (Das et al. 1996) indicates that the region between

motifs three and four is located far from the catalytic site, while motif six is located in and around the catalytic channel. To better view the relationship between FEM-2's pattern of sequence conservation and the three-dimensional structure of the PP2C domain, the three-dimensional structure was colored according to the level of conservation among all three FEM-2 orthologs (Figure 16A-D). The results suggest that there is a high concentration of conserved sequences on the catalytic face of the protein. When the same structure is colored according to the conservation among the *C. elegans* PP2C paralogs, it is apparent that the catalytic region differs between them (Figure 16E-H).

3.5 Foreign *fem-2* genes can rescue somatic feminization in *C. elegans* mutants

To test whether the sequence divergence among the FEM-2 proteins affects regions of the protein important for masculinizing activity in the soma a variety of constructs were introduced into *C. elegans fem-2(null)* mutants. Wild-type XO animals normally develop as males, while XO m-z- *Ce-fem-2(null)* animals are feminized in both the germ line and soma. The degree of somatic feminization is temperature dependent. At 25°C the mutant XO worms are completely female, while at 20°C they have an intersexual soma that includes a partially formed vulva, a hermaphroditic double-armed gonad, and an incomplete male tail (Hodgkin 1986). XO m-z- *Ce-fem-2(null)* animals, either carrying a construct array or untransformed, were examined for somatic phenotype. At 20° transgenic *Ce-fem-2*, *Cb-fem-2* or *Cs-fem-2* was able to rescue the somatic feminization in *C. elegans fem-2* mutants, as were chimeric constructs consisting of the *Ce-fem-2* promoter and UTRs fused to the *Cb-fem-2* or *Cs-fem-2* coding regions. In each case, transformed XO animals were obtained that

lacked a vulva, had a single-armed gonad, and a distinct male tail, while untransformed XO animals always had a vulva, a double-armed gonad, and a slightly masculinized tail (Figure 17; Table 7). The masculinization of the tail was not complete in any of the transgenic animals, including those carrying a *Ce-fem-2* transgene. At 25°C many transgenic animals were obtained for each construct that lacked a vulva, had a single-armed gonad, and a distinct male tail, while control animals were always completely female (data not shown).

3.6 Foreign *fem-2* genes cannot rescue germ-line feminization in *C. elegans* mutants

Genetic evidence indicates that *fem-2* has distinct targets in the germ line and soma in *C. elegans* (Hodgkin 1986), raising the possibility that the FEM-2 protein from another species could act on the targets in one tissue but not the other. To test whether transgenic *fem-2* from *C. briggsae* or CB5161 can rescue germ-line feminization in *C. elegans* a variety of constructs were injected into *fem-2(b245)* animals, which are unable to produce sperm when raised at 25°C. The advantage of using the *b245* mutation is that animals raised at 20°C can be injected and moved to 25°C. The progeny of the injected animals can then be examined for germ-line rescue by their ability to reproduce at 25°C. Assaying the direct progeny of injected animals is important because transgenes in *C. elegans* show progressive germ-line silencing over generations (Kelly et al. 1997). Transgenic *Ce-fem-2* was able to restore sperm production in *fem-2(b245)* animals, as indicated by the presence of hermaphrodites with eggs after growth at the restrictive temperature. The brood size of the rescued hermaphrodites was variable (between 15 and 188), and the average brood size was

106. *Cb-fem-2* or *Cs-fem-2* were unable to rescue the germ-line feminization, as were chimeric constructs consisting of the *Ce-fem-2* promoter and UTRs fused to *the Cb-fem-2* or *Cs-fem-2* coding regions (Table 8).

3.7 *fem-2(RNAi)* does not feminize *C. briggsae* hermaphrodites or CB5161 males

RNA-mediated interference (RNAi) was used to address whether *Cb-fem-2* or *Cs-fem-2* are sex-determining genes in their respective species. This technique has been used to silence genes in *C. elegans* (Fire et al. 1998), *C. briggsae* (Kuwabara 1996b; Streit et al. 1999; Molin et al. 2000; Rudel and Kimble 2001), and *C. remanei* (Haag and Kimble 2000; Rudel and Kimble 2001). RNAi against *fem-2* in *C. elegans* caused germ-line feminization in males and hermaphrodites, but did not feminize the male soma (Figure 18; Table 9). In *C. briggsae*, *fem-2(RNAi)* feminized the male germ line (Figure 18; Table 9), but not the hermaphrodite germ line. No reduction in the brood size of hermaphrodites arising from injected animals was detected, suggesting that normal numbers of sperm were produced (data not shown). Although RNAi did not shift the sex ratio towards females, a few intersex animals were observed, which had a male tail, a single-armed gonad, and what appeared to be a hermaphrodite vulva. However, the role of *Cb-fem-2* in the soma remains uncertain because few animals were affected, and the tail and somatic gonad were not feminized. In CB5161, *fem-2(RNAi)* did not cause noticeable changes, with males having a normal germ line and soma. No other obvious defects were observed in any of the species. These results suggest that *fem-2* may not play a significant role in the germ line of CB5161 males or *C. briggsae* hermaphrodites, or that RNAi is less effective at inhibiting *fem-2* activity in these species.

3.8 Polymorphisms in *fem-2* and F25D1.1

One test used to determine whether positive selection is driving the evolution of a protein involves isolating the same gene from two species and comparing the amount of amino acid sequence divergence (K_a) to the amount of silent base divergence (K_s) (for review see Kreitman and Comeron 1999). If a sequence is under no selective constraints, a pseudogene for example, K_a and K_s should nearly be equal. For most proteins K_a is much less than K_s, implying that purifying selection has been operating to preserve the amino acid sequence. When K_a is greater than K_s, selection for amino acid replacements (positive selection) is thought to have occurred. However, many instances of positive selection may be missed by the $K_a > K_s$ criterion. If substantial parts of a protein are subjected to purifying selection, K_a will not exceed K_s despite the occurrence of adaptive evolution at other sites. Also, if a protein is subjected to positive selection for a short period of time, continuous synonymous divergence will eventually make K_s greater than K_a. The McDonald-Kreitman test can be used to detect positive selection when there is not an excess of nonsynonymous divergence relative to synonymous divergence (McDonald and Kreitman 1991). It makes use of polymorphism measurements—the amount of variation in a gene within a species. The test assumes that if a gene is evolving neutrally, the ratio of replacement to synonymous differences within a species should equal the ratio between species. Under positive selection, the between-species ratio is expected to be greater. The high level of synonymous divergence between *C. elegans* and CB5161 (Table 6) makes it unlikely for any gene to exhibit a K_a value that is greater than K_s. To determine whether polymorphisms exist in *fem-2*, which could be used in the McDonald-

Kreitman test, portions of the gene were sequenced from various *C. elegans* isolates (Figure 19A,B). A segment of the *fem-2* paralog F25D1.1, which includes an intron, was also sequenced (Figure 19C). No differences were observed between the sequences obtained and the existing reference sequences from *C. elegans* strain N2. Although only small portions of the genes were sequenced, the absence of any differences suggests that the *C. elegans* population may be too homogenous for useful polymorphism data to be obtained. Consistent with this view are the results of a previous study of the calmodulin gene, in which no differences were found among eleven strains (Thomas and Wilson 1991).

3.9 Yeast two-hybrid screen for proteins that interact with FEM-2

Yeast two-hybrid screens have identified numerous protein interactions thought to be biologically relevant, including four believed to regulate sex determination in *C. elegans* (Chin-Sang and Spence 1996; Mehra et al. 1999; Clifford et al. 2000; Lum et al. 2000; Wang and Kimble 2001). Although *fem-2* is clearly required for specification of the male fate, its specific role is unclear, as its targets are not known. To identify substrates and proteins that might modulate the activity of FEM-2, a yeast two-hybrid screen was performed. Preparation of a suitable bait plasmid proved to be problematic, as a fusion containing the complete *fem-2* cDNA activated the reporter genes meant to signal an interaction. The same auto-activation was encountered with the *b245* allele of *fem-2*, despite its inability to generate a functional phosphatase. A construct consisting of the amino-terminal domain region of *fem-2* did not cause reporter gene expression on its own. Although certain interactions may be missed, a screen using this construct could be informative, as the amino-terminal domain of

FEM-2 is required for sex-determining activity. Approximately 3 x 10^5 clones from a *C. elegans* cDNA library were transformed into yeast bearing the *fem-2* bait plasmid. 37 colonies grew on plates that selected for expression of *HIS3* and *ADE2* reporter genes under the control of GAL binding sites. 27 of these were found to interact with a non-*fem-2* bait plasmid and were discarded. The remaining positive prey plasmids were retransformed into the original bait strain, and those that yielded colony growth were sequenced (Table 10). Six contain inserts that correspond to predicted *C. elegans* genes, and all are in the correct reading frame for production of a GAL4AD fusion protein. Two of the positives appear to encode a C-type lectin, which is a carbohydrate-binding protein known primarily as a component of the vertebrate immune system (Weis et al. 1998). One positive shows similarity to vacuolar ATPases, which are ATP-dependent proton pumps that acidify intracellular compartments in eukaryotic cells (Forgac 1999). Another encodes a portion of CUT-2, a secreted protein that forms part of the *C. elegans* cuticle (Lassandro et al. 1994). The remaining two positives do not resemble any proteins of known function.

3.10 Construction and expression of *fem-2* reporter genes

Determining FEM-2's precise role in *C. elegans* will require knowledge of its temporal and spatial expression patterns, both in males and hermaphrodites. Protein expression is most easily studied using GFP (green fluorescent protein) fusions, since living animals can be examined without any special preparation. If the fusion is introduced into a mutant strain lacking the protein of interest, phenotypic rescue may be observed. Rescue indicates that the fusion is functioning in the cells that require the endogenous protein. Cautious interpretation of transgene expression is required

however, because exogenous DNA usually forms multi-copy arrays in *C. elegans*, potentially leading to overexpression or misexpression. Thus only a subset of the cells expressing a rescuing fusion may really require it. Also, the arrays are usually extrachromosomal, meaning that results from animal to animal can vary because of random loss of the array. A new technique for introducing DNA into *C. elegans* may solve these problems, by mediating single-copy and low-copy chromosomal insertions (Praitis et al. 2001).

Previous attempts to make a functional GFP::*fem*-2 reporter by fusing the GFP coding sequence upstream or downstream of *fem-2* were unsuccessful (Dave Hansen, personal communication). One explanation for the lack of rescue is that the additional protein hinders normal folding of FEM-2. Another is that the reporter sequence blocks access to a portion of FEM-2 that mediates an interaction with another protein. Analysis of the structure and evolution of the PP2C domain revealed a region that is poorly conserved in sequence and size, and located away from the catalytic face (Figure 20). A site in the corresponding region of FEM-2 was chosen as the insertion point for a variety of reporter sequences, the rationale being that it may be located in a part of the protein that is not functionally important. Three separate constructs were prepared: one containing the GFP coding sequence, one containing the β-GAL coding sequence, and one containing three adjacent copies of the HA (hemagglutinin) coding sequence. The fusions, and the *in vitro* mutagenized version of *fem-2* that was used in their construction (pDP#PS1p.644), were separately injected into *fem-2(b245)* animals, to see whether they could restore fertility. Hermaphrodites transformed with pDP#PS1p.644 or the HA construct were able to produce progeny, while those

containing the *lacZ* or GFP fusion were sterile (Table 11). Not only were the *lacZ* and GFP fusions unable to rescue mutant worms, they dominantly feminized N2 hermaphrodite and male worms. The effect was progressive, as the F1 progeny of injected N2 animals appeared to be normal, but within a few generations all transgenic worms were feminized. A line of *fem-2(b245)* worms transmitting the HA construct was maintained at 25°C for several generations. However, the worms produced very few progeny per generation, indicating that the transgene was silenced, or that the fusion protein has low activity.

Although a FEM-2 fusion capable of fully rescuing mutant worms was not obtained, expression of the constructs was examined. Those that are expressed to some degree may be useful in future studies, possibly as single-copy integrants. The GFP and β-GAL fusion proteins were visualized in the dominantly-feminized lines of N2 worms, in individuals from the first few generations. GFP was typically observed in the pharynx, intestine, spermathecae, and vulva (Figure 21). β-GAL staining was only observed in the pharynx (data not shown). Expression of HA was detected in protein extracts from transgenic *fem-2(b245)* worms, using an anti-HA antibody (Figure 22). Antisera raised against FEM-2 binds with what appears to be the same protein, as well as the endogenous FEM-2 protein. Although the antisera reacts specifically on western blots, attempts to visualize FEM-2 in fixed animals have not been successful. Control animals that should lack or have reduced levels of FEM-2 always show the same staining pattern as wild-type animals (data not shown).

3.11 Characterization of a human homolog of *fem-2*

The protein that most closely resembles FEM-2, apart from its *C. briggsae* and CB5161 orthologs, is a predicted human PP2C (encoded by the KIAA0015 cDNA) identified by Nomura et al. (1994) (Figure 23). This similarity raised the possibility that the human protein plays a related role in development. To better understand the function of KIAA0015, steps were undertaken to map the corresponding genomic copy, determine its sequence, and examine its expression. The gene was mapped to bin 8.2 (Marcia Budarf, University of Pennsylvania), a region within 22q11.2 delineated by a somatic cell hybrid panel (Budarf et al. 1996). A coding portion of the cDNA was then used to isolate genomic clones from a chromosome 22-specific cosmid library. Ten positives were obtained, and Southern analysis using a second probe was performed to verify that they contained sequence found in KIAA0015 (Figure 24). A megabase of sequence spanning the immunoglobulin λ locus was published a short time later, including the positive cosmid 109g12 (Kazuhiko et al. 1997). The location of this clone in the λ locus contig is consistent with its previously determined map position, and its sequence confirms that the KIAA0015 gene is present. Northern analysis revealed that the gene is widely expressed, and possibly alternatively spliced (Figure 25). The additional bands of hybridization could also be the result of RNA degradation, or hybridization to the products of another gene. The mapping and expression data do not rule out a sex-determining function, but could suggest that KIAA0015 has a more general role not related to sex determination. Recently, the rat ortholog of this gene was isolated as CaM-kinase phosphatase, a protein that dephosphorylates and deactivates calmodulin-dependent protein kinase II

(CaM-kinase II). CaM-kinase phosphatase is present in all tissues, and is presumed to regulate intracellular calcium signaling (Kitani et al. 1999).

3.12 Characterization of a zebrafish pyruvate dehydrogenase phosphatase

One of the sequences obtained from zebrafish using degenerate-oligo PCR is predicted to encode a pyruvate dehydrogenase phosphatase (PDP), a protein that dephosphorylates and activates the pyruvate dehydrogenase complex in mitochondria (Lawson et al. 1993; Huang et al. 1998). Sequence comparisons place the PDP proteins into a distinct subfamily within the PP2C family, separate from FEM-2 and its paralogs (Bork et al. 1996). For this reason, the PDP-encoding clone was excluded from the evolutionary analysis of FEM-2. Angela Manning (University of Alberta) decided to examine the expression of the zebrafish gene, in part because there were few suitable sequences available for testing zebrafish northern blots. Surprisingly, expression of the partial *pdp* sequence was detectable only in females (Figure 26). This result renewed our interest in the gene, which we then studied on several fronts. A larger portion of the *pdp* coding sequence was isolated from a zebrafish cDNA library, allowing more meaningful comparisons to be made. The predicted protein contains all the requisite metal and phosphate-binding residues, and shows similarity to known PDP proteins outside of the PP2C domain (Figure 27). When it is aligned with the two mammalian isoforms separately, the similarity is higher for PDP2 than PDP1 (53% identity and 65% similarity vs. 47% identity and 59% similarity, respectively). Based on this comparison we refer to the zebrafish gene as *pdp2*. The genomic copy of *pdp2* was mapped by PCR using a recently prepared radiation hybrid

panel (Hukriede et al. 1999). The results place the gene on linkage group 18, 2.33 cR from the marker Z8488.

Why would a protein that regulates carbohydrate metabolism be expressed at higher levels in females? Initially we wondered if PDP had acquired a role in sex determination or differentiation. However, there were less intriguing possibilities. A significant proportion of the adult female zebrafish consists of oocytes. One explanation for the apparent sex-specificity of *pdp2* is that transcripts are supplied maternally. Northern analysis suggests that this indeed the case, as RNA from oocytes accounts for most if not all of the transcripts observed in females (Figure 28).

4 Discussion

4.1 PP2Cs in CB5161

Protein phosphatases play important roles in regulating a variety of cellular processes (Barford 1996; Pilgrim et al. 1995). The PP2C class of phosphatases is one of the two classes of enzymes that can dephosphorylate serine and threonine residues (Bork et al. 1996), and they are defined biochemically as having phosphatase activity that is Mg^{2+}-dependent or Mn^{2+}-dependent and okadaic acid resistant (Cohen 1989). Sequence analysis of the phosphatases meeting these criteria shows that they comprise a conserved protein superfamily (Bork et al. 1996). In addition to the FEM-2 sex-determining protein, there are three predicted PP2C genes in *C. elegans*, two of which contain a PP2C domain that is close in size to the one found in FEM-2. From a variety of sequence comparisons it appears that the orthologs of the *C. elegans fem-2* gene were isolated from *C. briggsae* (Hansen and Pilgrim 1998) and CB5161. Cb-FEM-2 and Cs-FEM-2 are much more similar to *C. elegans* FEM-2 than to other *C. elegans* PP2Cs, they contain conserved sequence segments absent from other PP2Cs, and the positions of three introns are identical in all three *fem-2* sequences. The grouping of the additional CB5161 PP2C sequences with their *C. elegans* orthologs was straightforward, based on their high level of sequence identity. It is possible that missing sequence information has led to incorrect ortholog assignments. Even with complete genome knowledge ortholog assignments remain uncertain because of the possibility of gene loss.

Because the genomic and cDNA sequences of *fem-2* have been obtained from three different species, the evolution of intron position can be explored with little

concern that incorrect exon/intron predictions have been made. Furthermore, the intron-containing regions of the *fem-2* genes can be unambiguously aligned, preventing perceived intron movement resulting from alignment choice. Changes in intron position and number have been proposed to arise from random loss of introns from an intron-rich ancestral gene (Crabtree et al. 1985). However, further gene characterization has revealed that loss alone cannot explain the gene structures observed today—the ancestral genes would have to contain far too many introns (Stoltzfus et al. 1997). Two alternative processes that could be involved are insertion and sliding. The latter has been proposed in cases where the number of introns and their general locations are conserved, but not the exact position of one or more introns (Brenner and Corrochano 1996; Jellie et al. 1996). There is little evidence for intron sliding, and published examples often turn out to be the result of ambiguous sequence alignments or sequencing errors (for review see Stoltzfus et al. 1997). In this study of the evolution of *fem-2*, divergence in intron position is apparent (Figure 12; Figure 13) and appears to result from intron loss or gain. With knowledge of the evolutionary relationships between the species (i.e. which two are most closely related), it should be possible to trace more precisely the specific events that led to current gene structures.

4.2 Cs-FEM-2 appears to be a functional PP2C

To assess how the observed evolutionary differences in FEM-2 might affect the function of the protein, the three-dimensional structure of human PP2Cα was examined. In other studies the availability of x-ray crystallographic structures has made it possible to identify evolutionary changes that are likely to have important functional consequences. For example, comparisons of plant chitinase sequences with

a chitinase three-dimensional structure revealed that numerous replacements occur in the active site cleft of the enzyme, consistent with coevolutionary interactions with pathogens (Bishop et al. 2000). In this analysis of FEM-2 the residues predicted to bind metal and phosphate ions were compared between species and found to be invariant. The conservation of the remaining residues appears to be related to their position in the PP2C domain, with those covering the catalytic face being conserved, and those remote from the catalytic site having diverged. The conservation of the catalytic face suggests that the protein dephosphorylates a substrate domain that is also similar among the three species examined. Currently, the targets of FEM-2's phosphatase activity in *C. elegans* and the other nematode species are not known.

4.3 Rapid evolution of FEM-2

Comparison of the *C. elegans* PP2Cs FEM-2, F25D1.1, and T23F11.1 with their orthologs in CB5161 shows that FEM-2's PP2C domain is the most diverged. A previous analysis of sequences suggests that FEM-2 is also evolving rapidly between *C. briggsae* and *C. elegans* (Hansen and Pilgrim 1998). Why is FEM-2 evolving more rapidly than other proteins? In cases where positive selection is thought to cause sequence divergence, the active sites of proteins are often affected. For example, positive selection is believed to have caused a high level of amino acid divergence in the antigen-binding clefts of certain major-histocompatibility-complex molecules (Hughes et al. 1990), the active center regions of serine protease inhibitors (Hill and Hastie 1987), and the active site clefts of chitinases (Bishop et al. 2000). Our comparison of FEM-2's PP2C domain divergence to the three-dimensional structure of another PP2C domain suggests that the active site region of FEM-2 is conserved.

Also, although FEM-2's PP2C domain contains more diverged residues than the other PP2C domains, the locations of many of the diverged residues are in regions that typically diverge between PP2C orthologs. Despite the consistency of the divergence pattern with neutral evolution, positive selection cannot be ruled out. Tests commonly used to distinguish between the two cannot be applied here because the high level of synonymous divergence between the sequences would mask evidence of positive selection, and polymorphisms within *C. elegans* are rare even in noncoding DNA (this work; Thomas and Wilson 1991; Kock et al. 2000). It is not clear why sex-determining genes would be subjected to positive selection. One hypothesis is that changes might facilitate reproductive isolation during the early stages of speciation, by causing hybrid sterility (Whitfield et al. 1993). Another explanation is that there could be selection for mutations that adjust the level or timing of gamete production towards an optimum (Hodgkin and Barnes 1991). Accelerated neutral evolution may occur because changes in sex-determining proteins are tolerated more than changes in other proteins. In *C. elegans* a variety of sex-determining gene mutations have a slight feminizing or masculinizing effect and yield fertile animals. Proteins that regulate other processes may not be as free to diverge, since changes in their activity may not simply skew the cell towards an alternate viable state. Thus equivalent mutations in two similar proteins, one that regulates sex determination and one that does not, could have dramatically different effects on the fitness of the organism.

4.4 Portions of FEM-2 important for germ-line function may not be conserved

Although the diverged sequences in FEM-2's PP2C domain are concentrated away from the catalytic site, they could still play an important role in binding a targeting

protein or substrate. Also, FEM-2 contains a long amino-terminal domain that is absent from most other PP2C proteins and required for sex-determining function in *C. elegans*, but not for phosphatase activity (Hansen and Pilgrim 1998). The *C. briggsae* and CB5161 *fem-2* orthologs were tested for activity in *C. elegans*, to address more directly if the protein divergence affects function. *Cb-fem-2* has been previously shown to rescue the somatic feminization that occurs in XO animals carrying a mutant form of *fem-2* (Hansen and Pilgrim 1998). The *C. briggsae* orthologs of *her-1* and *tra-1* have also been introduced into *C. elegans*, and in both cases the genes were found to function in many of tissues regulated by the *C. elegans* gene (de Bono and Hodgkin 1996; Streit et al. 1999). In the case of *tra-1*, no rescue was observed in the somatic gonad or the germ line (de Bono and Hodgkin 1996). Although the *Cs-fem-2* and *Cb-fem-2* transgenes tested in this study could rescue the somatic feminization, they were unable to rescue the germ-line feminization caused by a *fem-2* mutation. This result suggests that the regions of FEM-2 that mediate an interaction with a germ-line substrate or targeting protein in *C. elegans* are not conserved. It seems unlikely that the absence of germ-line rescue reflects a lack of transgene expression, since a *Ce-fem-2* transgene was able to function robustly in the same assay. Even when the *fem-2* coding sequences from the other species were surrounded by the *Ce-fem-2* regulatory regions germ-line rescue was not observed.

Although the ability of the foreign *fem-2* genes to function in the soma but not the germ line of *C. elegans* could reflect changes in the germ-line activity of the FEM-2 proteins relative to their somatic activity, another possibility exists. They could act equally well in both tissues, and the lack of rescue in the germ line could be the

consequence of lower levels of transgene expression in the germ line relative to the soma.

4.5 The germ-line role of *fem-2* in *C. briggsae* and CB5161

The fact that a gene from a foreign species can rescue a mutant phenotype does not show that the gene regulates the same pathway in both species. Recent work in Dipteran insects suggests that some orthologous genes may have a sex-determining role in one species but not another. In *Drosophila* the *Sex-lethal* (*Sxl*) gene is a key regulator of sex determination, and clear orthologs have been isolated from *Chrysomya* (Muller-Holtkamp 1995), *Megaselia* (Sievert et al. 1997), *Musca* (Meise et al. 1998) and *Ceratitis* (Saccone et al. 1998). However, the alternative splicing of *Sxl* that determines sex in *Drosophila* is not observed in any of the other species, suggesting that its developmental role may not be conserved. The most downstream *Drosophila* gene is *doublesex*, and its ortholog has been isolated from *Megaselia*, (Sievert et al. 1997) and *Bactrocera* (Shearman and Frommer 1998) and in both cases it shows the sex-specific splicing observed in *Drosophila*. Although the function of the orthologs remains to be tested in the other species, the results thus far are consistent with the model proposed by Wilkins (1995), in which sex-determining pathways are built from the bottom up. In *Caenorhabditis* the technique of RNA-mediated interference can be used to silence specific genes, and it has been used to test whether specific sex-determining genes have a conserved sex-determining role. RNAi has been performed against *Cb-tra-2* and *Cb-her-1*, and in both cases the phenotype in *C. briggsae* was similar to the *C. elegans* phenotype (Kuwabara 1996b; Streit et al. 1999). The *tra-2* gene was isolated from the closely related male/female species *C.*

remanei and RNAi performed against *Cr-tra-2* showed that *tra-2* has a conserved sex-determining role (Haag and Kimble 2000). In this study RNAi against *Ce-fem-2* caused the germ-line feminization seen in animals carrying a *fem-2* mutation. When RNAi was performed against the *fem-2* ortholog in *C. briggsae*, the male germ line was feminized, but not the hermaphrodite germ line. There are a variety of explanations for why *fem-2(RNAi)* in *C. briggsae* appears to feminize the male germ line but not the hermaphrodite germ line. RNAi may act more effectively at the time or place that *fem-2* activity is required in the male, the male germ line may be more sensitive to a reduction in *fem-2* activity, or slight germ-line feminization may be easier to observe in male animals. A more intriguing possibility is that *fem-2* in *C. briggsae* regulates germ-line sex in males but not hermaphrodites. *Cb-fem-2* not having a role in the hermaphrodite germ line is unexpected, given that the other sex-determining gene orthologs that had been examined in *C. briggsae* appear to have the same role as their *C. elegans* counterparts. Furthermore, the previously tested genes act upstream of *fem-2*. According to the model proposed by Wilkins (1995), the pathway has grown from the bottom up, suggesting that any genes acting downstream of a component with a conserved role should themselves have a conserved role. The absence of a *fem-2(RNAi)* phenotype in CB5161 is more difficult to interpret, because no cases of successful RNAi have been reported for this species. However, this observation does raise the possibility that *fem-2* is not important for sperm production in CB5161 males.

4.6 Emerging results from studies of *Cb-fem-1* and *Cb-fem-3*

Given the unexpected RNAi phenotypes for the *fem-2* orthologs, a lingering question has been whether or not feminization will be observed when RNAi is performed against orthologs of the other *fem* genes. The common position of the *fem*s in the genetic hierarchy, and the interaction between FEM-2 and FEM-3, suggest that all three proteins may act as a complex. A reasonable prediction is that in a cell that does not require *fem-2* for male development, *fem-1* and *fem-3* will also be dispensable. *Cb-fem-1* and *Cb-fem-3* have recently been isolated (Jeb Gaudet personal communication; Eric Haag personal communication), and the effects of RNAi against these genes have been examined. The results thus far are consistent with *fem-1* and *fem-3* having no germ-line role, at least in hermaphrodites. The RNAi phenotypes for the *fem* genes could covary between species for reasons other than gene product interdependence. Because gene silencing by dsRNA depends on its own genetic pathway, RNAi differences could just as easily reflect divergence of the RNAi-mediating machinery. Mutations in the *C. briggsae fems* will have to be obtained for the roles of these genes to be described more convincingly.

4.7 How might interspecies differences between pathways arise?

There are a variety of explanations for how the FEM proteins could end up with different roles in *C. briggsae* and *C. elegans*. The common ancestor of these species may have had a sex-determining pathway that resembles that of *C. elegans*, with changes in *C. briggsae* subsequently making the FEMs dispensable in the hermaphrodite germ line. Such changes would have to bypass the need for the FEM proteins, by incorporating new sex-determining genes, or by modifying existing ones. Whether these changes occur because there is pressure to change or change is tolerated

is not clear. Another possibility is that the ancestral species did not use the FEMs for germ-line sex determination, and that they were incorporated into the *C. elegans* but not the *C. briggsae* pathway. A limitless number of less parsimonious routes could have also led to the differences.

The model proposed by Wilkins (1995), in which selection causes the sex-determining pathway to grow from the bottom up, could still be consistent with the FEMs having a reduced role in *C. briggsae*. Instances have been reported of established pathways diverging over time without any accompanying change in the phenotypic outcome (True and Haag 2001). This phenomenon, referred to as developmental system drift (DSD), could obscure supporting evidence for bottom up growth by randomly introducing and removing components at any level of a pathway. Regardless of how interspecies differences in pathways arise, it will be important to determine how common these differences are, as this information could be used to estimate the general importance of data obtained in model organisms.

The prevailing view held by molecular biologists is that if a gene in one species has a clear ortholog in another, information about the biological function of one can be applied with confidence to the other. This work and the findings of others investigating the evolution of sex determination suggest that this view may be incorrect, at least some of the time. It remains to be determined if evolutionary change of gene function between closely related species is a common occurrence. Few other pathways have been studied in the detail necessary to reveal differences, most likely because pathway conservation is assumed because of sequence conservation. Also, the studies that are done often use a questionable method to measure pathway

relatedness. The experiment involves expressing a gene from one species in the cells of another. If the foreign gene can substitute for the endogenous gene (i.e. the effect is the same as when the endogenous gene is expressed), the foreign gene is said to have the same biological function. An analogy for this type of experiment is the following: a belt that connects an electric motor to a cooling fan breaks, so you look for a replacement belt in a box of miscellaneous parts. You select the one that most closely resembles the one you want to replace. If the replacement belt works, then it must have originally been used to connect an electric motor to a cooling fan (which might not be the case). For example, Zarkower (2001) presents the following as evidence that the homolog of *fem-1* is sex-determining in mammals: "Genes related to *fem-1*, called *Fem1a* and *Fem1b*, have been identified in mice, and *Fem1a* can very weakly rescue *C. elegans fem-1* mutants." The opposite is also assumed, that is, if the most similar part does not work, then it did not connect a cooling fan to a motor. For example, Tan et al. (2001) say the following: "Does hFEM-2 have a role in sex determination? This possibility seems remote as *C. briggsae fem-2*, which has 45% amino acid identity to *C. elegans fem-2*, can only weakly rescue the FEM-2 phenotype in *C. elegans*." Because the contexts that Cb-FEM-2 and Ce-FEM-2 operate in are different, expression of one in the new context is really just a measure of its ability to act as a molecular substitute, usually in an overexpressed form, and says nothing about the pathways it normally regulates. Ce-TRA-2 and Cb-TRA-2 are not interchangeable—Ce-TRA-2 interacts with Ce-TRA-1, and Cb-TRA-2 interacts with Cb-TRA-1, but no interactions occur across species (Wang and Kimble 2001).

Nevertheless, *Ce-tra-2* and *Cb-tra-2* mutants have the same sex-related phenotypes (Kuwabara 1996b).

4.8 Function of *fem-2* orthologs in more distant species

The molecular characterization of FEM-2 raised the possibility that knowledge of its role could be applied to its closest human relative, KIAA0015. According to the results of this study, KIAA0015 is widely expressed and located in a genomic region that has not been associated with sex determination. Given these results, how might the roles of FEM-2 and KIAA0015 be related? At this point it is important to distinguish between biological function and molecular function. If we assume that KIAA0015 does not regulate sex determination, it and FEM-2 perform different biological functions. Based on the presence of the required metal binding and phosphate-binding residues, it is reasonable to assume that the human protein acts as a protein phosphatase. Thus KIAA0015 and FEM-2 likely perform the same molecular function to some extent—they dephosphorylate protein substrates. The similarity of their molecular function could be more extensive, with the two proteins acting on related substrates—these substrates, or any of the targets of the substrates, could be the ones that acquired a different molecular function that in turn changed the biological function of the phosphatases that regulate them. Thus some information gained about the molecular interactions between FEM-2 and its targets may be applicable to KIAA0015 and vice versa.

Recently, the rat ortholog of KIAA0015 was isolated in a search for CaMKPase, the phosphatase that acts on CaM kinase II (Ca^{2+}/calmodulin-dependent kinase II) (Kitani et al. 1999). CaM kinases, which are activated by phosphorylation,

regulate a variety of processes including muscle contraction, neurotransmitter release, gene expression, cell proliferation, and apoptosis (Nairn and Picciotto 1994; Wright et al. 1997). CaMKPase was originally purified from rat brain using a synthetic peptide corresponding to the autophosphorylation site of CaM kinase II (Ishida et al. 1998a). The protein was then confirmed to dephosphorylate activated CaM kinase I, II, and IV (Ishida et al. 1998a; Ishida et al. 1998b). Identification of the corresponding cDNA revealed that the protein is 79% identical to KIAA0015 (Kitani et al. 1999). A group specifically searching for a human homolog of *fem-2* has re-isolated the KIAA0015 cDNA (Tan et al. 2001). Their interest in *fem-2* stemmed from previous work in which they demonstrated that a human protein called F1Aα, which resembles *C. elegans* FEM-1, promotes apoptosis in mammalian cells (Chan et al. 2000). They wondered if KIAA0015 also regulates apoptosis, possibly by interacting with F1Aα. Because the rat ortholog had been shown to function as a CaMKPase, they tested KIAA0015 for this activity. As expected, the protein dephosphorylates phosphorylated CaM kinase II (the other CaM kinases were not examined as substrates). These results raise the possibility that FEM-2 regulates sex determination by dephosphorylating a CaM kinase that has acquired a sex-determining role relatively recently in nematode evolution, as a direct or indirect regulator of TRA-1 for example. This sex-determining kinase may be essential for other pathways, such that it cannot be identified using screens for inappropriately feminized or masculinized worms. FEM-2 could also dephosphorylate CaM kinases that regulate other processes. Mutations in *fem-2* may appear to only cause feminization because other PP2Cs can

substitute for FEM-2 to some degree, or special conditions are required to observe the other phenotypes.

In addition to possessing CaMKPase activity, KIAA0015 was shown to promote apoptosis (Tan et al. 2001). The FEM proteins can be said to regulate apoptosis in *C. elegans* to some extent, because the terminal regulator of the somatic sex determining pathway, TRA-1, regulates the apoptosis-promoting gene *egl-1* in two neurons required for hermaphrodite egg laying (Conradt and Horvitz 1999). In hermaphrodites, the TRA-1 protein represses *egl-1* transcription by binding to its promoter, and the two neurons survive. In males TRA-1 is not active, the EGL-1 protein is expressed, and the neurons die. The effect of EGL-1 on apoptosis is mediated by an interaction with CED-9. This interaction prevents CED-9 from inhibiting CED-4, such that CED-4 is able to activate the CED-3 caspase (del Peso et al. 2000). The effect of the FEMs on the egg-laying neurons occurs completely through TRA-1—if TRA-1 activity is removed by mutation, the FEMs are no longer needed for apoptosis to occur. Also, the FEMs are not thought to play a general role in promoting apoptosis, although a redundant role cannot be excluded. In mammals, F1Aα binds directly to the CED-4 homolog Apaf-1, and F1Aα is cleaved by the CED-3 homolog caspase-3 (Chan et al. 2000). KIAA0015, in addition to acting as a CaMKPase, interacts directly with F1Aα (Tan et al. 2001). It is not clear whether KIAA0015 promotes apoptosis by regulating interactions between F1Aα and CED-3, or exclusively by dephosphorylating CaM kinases. Based on the results obtained thus far, it appears that the apoptotic roles of the *C. elegans* FEMs and mammalian orthologs were acquired independently, as the mechanisms they use to control the

apoptotic machinery are distinct. The finding that the FEM-1 and FEM-2 homologs in mammals interact and induce apoptosis is interesting from an evolutionary standpoint, because it suggests that some components of the *C. elegans* sex-determining pathway may operate together in other species, but in pathways that regulate other processes.

To date the only reported case in which the sequence and biological function of a sex-determining gene is conserved between *C. elegans* and mammals involves *mab-3*, a downstream target of *tra-1*. In *C. elegans*, *mab-3* controls the differentiation of sex-specific neuroblasts and regulates the transcription of yolk protein genes (Shen and Hodgkin 1988). Its protein product contains a DM domain—a type of zinc-finger DNA-binding motif (Raymond et al. 1998). A human homolog of *mab-3*, called *DMT1*, maps to a region of chromosome 9 implicated in human XY sex reversal and is expressed only in testis (Raymond et al. 1998). In mice the same gene is required in males for testis differentiation (Raymond et al. 2000). A caveat of these findings, in terms of what they say about the evolution of sex determination, is that there are five human genes and four *C. elegans* genes that are similar to *mab-3*. *DMT1* may not be the ortholog of *mab-3*, in which case their involvement in sexual development simply represents use of a similar molecular component rather the evolutionary conservation of a pathway.

4.9 Future directions

4.9.1 Identifying FEM-2's targets and partners

Yeast two-hybrid screens are a popular means of identifying protein-protein interactions, including those involving PP2Cs (Welihinda et al. 1998; Strovel et al. 2000; Leung-Hagesteijn et al. 2001). However, they tend to pick up positives that do

not reflect biologically relevant interactions, and some of these cannot be eliminated using control tests. For this reason, the results of a yeast two-hybrid screen need to be interpreted with caution. None of the positives obtained using FEM-2 as bait correspond to known sex-determining genes, making it impossible to assess their validity offhand. RNAi is one technique that could be used to explore the relevance of the results. A reasonable prediction is that inappropriate feminization or masculinization will occur if a FEM-2 substrate or accessory gene is silenced. RNAi data for some of the positives may soon be available from large-scale projects aimed at testing every gene in the genome. GFP-promoter fusions could also be useful, as they can be used to assess whether or not a protein is expressed in the right place and at the right time for an interaction to occur. Ultimately genetic evidence is the most convincing, and trying to support a two-hybrid result without it may be difficult. It may be possible to assign some positives to known mutations, using the detailed physical and genetic maps of the *C. elegans* genome. Alternatively, deletions could be obtained using the PCR-based selection schemes that have been developed. In either case the phenotype associated with the mutation may be informative on its own, or in combination with mutations in other genes that regulate sex determination.

A screen in which a FEM-3 fusion was used as bait yielded two complete *fem-2* cDNAs (Chin-Sang and Spence 1996). Thus the absence of *fem-3* from the list of positives obtained in this study is somewhat unexpected. There are several explanations for why a *fem-3* clone was not isolated. A cDNA in the correct reading frame may not have been included in the subset of screened cDNAs, the partial FEM-2 fusion used may not be sufficient for the interaction between the proteins, or the

proteins may not interact as well with the two-hybrid fusion domains swapped. Another larger-scale screen using a more complete FEM-2 fusion could address the first two issues. The FEM-2 protein contains a stretch of negatively charged residues near its carboxyl end. Removal of this part of the protein could potentially alleviate the reporter expression problem encountered with full-length FEM-2, as acidic domains stimulate transcription in yeast (Struhl 1995).

Information about the substrates of other PP2Cs could potentially be used to identify FEM-2 substrates. Many known PP2C targets are kinases whose activity is regulated by phosphorylation (Table 2). The intensity with which these kinases are studied is likely the reason for their being the most common type of substrate at present. Other targets include an ion channel (Travis et al. 1997), a protein that regulates pre-mRNA splicing (Murray et al. 1999), a cytoskeleton component (Hishiya et al. 1999), and an inhibitor of *Wnt* signaling (Strovel et al. 2000). Although none resemble a sex-determining protein in *C. elegans*, sequence comparisons could still be informative. For example, PP2Cs may tend to recognize phosphorylated residues in a particular sequence context. As the specific residues that are dephosphorylated by PP2Cs are identified, *C. elegans* proteins can be examined for similar sequences as a way of identifying candidate substrates. Perhaps one of the known substrates does have a sex-determining role in the worm. Its absence from the current pathway could result from it having an essential role in other aspects of development, such that mutations in the corresponding gene are lethal. Future proteins identified by yeast two-hybrid screens or other methods may be good candidates for bona fide FEM-2 targets if they resemble a known PP2C substrate.

4.9.2 Determining when and where FEM-2 is expressed

Future work should also attempt to more accurately determine the temporal and spatial expression pattern of FEM-2 in both males and hermaphrodites. Antisera will likely be necessary for this analysis, given the problems associated FEM-2::GFP fusions. An advantage of using antisera is that mutant worms can be examined without the hassle of introducing constructs by crosses or injection. The localization and expression of FEM-2 in the absence of FEM-1 and FEM-3 could be particularly informative, as these three proteins may act as a complex, in which some members target others to specific parts of the cell. Existing FEM-2 antisera raised against the entire protein (Petra Jackle, University of Alberta) works well on western blots but not with fixed animals. New antisera, raised against a smaller segment of the protein, might yield better whole animal staining. Some considerations when choosing a segment to use will include its location on the three-dimensional structure, its conservation among PP2Cs, and its conservation among FEM-2 orthologs. The use of a FEM-2 portion that is conserved among *C. elegans*, *C. briggsae*, and CB5161 may facilitate interspecies comparisons of expression.

4.9.3 Isolating mutations in *C. briggsae* sex-determining genes

RNAi is often touted as a powerful tool for investigating gene function, and in many cases it has proven to be useful. Some problems are associated with this technique, as certain tissues seem to be resistant to RNA-mediated silencing, and phenotypes are often not as severe or penetrant as those obtained with a corresponding null mutation. A method for introducing mutations into specific genes in *C. elegans* has not been developed. However, PCR-based screens for worms carrying deletions in genes of

interest have been productive. The sequences will soon be available for the *C. briggsae* orthologs of six *C. elegans* sex-determining genes. Using this information, a PCR screen for deletions in *C. briggsae* could be performed. Mutant worms could then be analyzed to determine whether the phenotypic effects resemble those observed in *C. elegans*. By comparing null mutants of one species to those of another, it should be possible to make more rigorous conclusions regarding the evolution of gene function.

A complementary approach would be to perform genetic screens in *C. briggsae*. These could be modeled after the ones used to decipher the sex-determining pathway in *C. elegans*. It is reasonable to assume that many of the mutations will be in orthologs of known components of the *C. elegans* pathway. The more intriguing mutations, if they are obtained, will be those that occur in genes with orthologs that have no known sex-determining role. These genes might be specific to the *C. briggsae* pathway, or the mutant phenotype in *C. elegans* might be more difficult to observe. In either case the results will be of interest.

Table 1. *C. elegans* sex-determining genes and proteins.

Gene	Loss-of-function phenotype	Protein domains, motifs, or signals	Protein functional information
fbf-1	XX excess sperm; XO normal	Related to RNA-binding proteins	Binds NOS-3
fbf-2	XX excess sperm; XO normal	Related to RNA-binding proteins	-
fem-1	XX fertile females; XO fertile females	Ankyrin repeats	Binds FEM-2
fem-2	XX fertile females; XO fertile females	Protein phosphatase type 2C	Has phosphatase activity; binds FEM-3
fem-3	XX fertile females; XO fertile females	Novel	Binds FEM-2; binds TRA-2
fog-1	XX oocytes only; XO oocytes only	CPEB (cytoplasmic polyadenylation element binding)	-
fog-2	XX oocytes only; XO normal	F-box	Binds GLD-1/*tra-2* mRNA complex
fog-3	XX oocytes only; XO oocytes only	Similar to vertebrate Tob, BTG1, and BTG2	-
fox-1	XX normal; suppresses XO-specific lethality and feminization caused by duplications of left end of X	RRM-type RNA-binding	Binds RNA
gld-1	XX oocytes only; XO normal	STAR RNA-binding	Binds many RNA targets, including *tra-2*
her-1	XX normal; XO form hermaphrodites	Secretory signal	Acts cell non-autonomously
laf-1	XX lethality and feminization; XO lethality and feminization in soma and germ line	Not yet cloned	-
mab-3	XX normal; XO abnormal tail and synthesizes yolk proteins	DM DNA-binding motif	Binds site in *vit-2* promoter
mog-1	XX sperm only; XO normal	DEAH-box RNA helicase	-
mog-2	XX sperm only; XO normal	Not yet cloned	-
mog-3	XX sperm only; XO normal	Not yet cloned	-

mog-4	XX sperm only; XO normal	DEAH-box RNA helicase	-
mog-5	XX sperm only; XO normal	DEAH-box RNA helicase	-
mog-6	XX sperm only; XO normal	Not yet cloned	-
nos-1	XX excess sperm; XO normal	Related to *Drosophila* Nanos	-
nos-2	XX excess sperm; XO normal	Related to *Drosophila* Nanos	-
nos-3	XX excess sperm; XO normal	Related to *Drosophila* Nanos	Binds FBF-1
sdc-1	XX weak masculinization; XO normal	Zinc fingers	-
sdc-2	XX masculinization and lethality; XO normal	Novel	Binds *her-1* promoter and localizes to X chromosomes
sdc-3	XX masculinization and lethality; XO normal	Zinc fingers	Localizes to X chromosomes
sex-1	XX show dosage compensation defects and masculinization; XO normal	Nuclear hormone receptor	Binds *xol-1* promoter
tra-1	XX low fertility males; XO low fertility males	Zinc fingers	Binds to DNA near *egl-1* and *mab-3*
tra-2	XX non-mating pseudomales; XO normal	Transmembrane domain	Binds FEM-3; Binds TRA-1
tra-3	XX sterile pseudomales; XO normal	Calpain protease	Cleaves TRA-2
xol-1	XX normal; XO die as embryos or small crumpled feminized L1 larvae	Novel	Binds *sdc-2* promoter

Table 2. PP2Cs and their putative substrates.

Organism	Enzyme[a]	Substrate[a]	Reference
Arabidopsis	KAPP	RLK5	Stone et al. 1994
Human	PP2Cα	CFTR	Travis et al. 1997
Human	PP2Cα	MKK6	Takekawa et al. 1998
Human	PP2Cα	p38	Takekawa et al. 1998
Human	PP2Cα	SEK	Takekawa et al. 1998
Budding yeast	Ptc2p	Ire1p	Welihinda et al. 1998
Budding yeast	Ptc2p	Cdc28p	Cheng et al. 1999
Budding yeast	Ptc3p	Cdc28p	Cheng et al. 1999
Human	PP2Cα	platelet moesin	Hishiya et al. 1999
Human	CaMKPase	CaMKII	Kitani et al. 1999
Human	PP2Cγ	Spliceosome component[b]	Murray et al. 1999
Hamster	PP2Cα	CFTR	Zhu et al. 1999
Human	PP2Cα	Cdk2	Cheng et al. 2000
Human	PP2Cβ	Cdk2	Cheng et al. 2000
Human	PP2Cα	Axin	Strovel et al. 2000
Human	ILKAP	ILK1	Leung-Hagesteijn et al. 2001

[a]Abbreviations: CaMKII, Ca^{2+}/Calmodulin-dependent protein kinase II; CaMKPase, Ca^{2+}/Calmodulin-dependent kinase phosphatase; Cdc, cell division cycle; Cdk, cyclin-dependent protein kinase; CFTR, cystic fibrosis transmembrane conductance regulator; ILK, integrin-linked kinase; ILKAP, integrin-linked kinase-associated phosphatase; KAPP, kinase-associated protein phosphatase; MKK, mitogen-activated protein kinase kinase; Ptc, phosphatase two C; RLK, serine-threonine receptor-like kinase; SEK, SAPK/ERK kinase; SAPK, stress-activated protein kinase; ERK, extracellular signal-regulated kinase.

[b]The precise target is not known but PP2Cγ is associated with the spliceosome and is required for spliceosome assembly.

Table 3. Phenotypes associated with mutations in PP2C genes.

Organism	PP2C gene[a]	Mutant phenotype	Reference
Arabidopsis	*ABI1*	Insensitive to abscisic acid	Leung et al. 1994
Budding yeast	*PTC1*	Growth defect[b]	Maeda et al. 1994
Fission yeast	*ptc1+*	Heat sensitive	Shiozaki and Russell 1994
C. elegans	*fem-2*	Feminized	Pilgrim et al. 1995
Fission yeast	*ptc1+*, *ptc2+*, *ptc3+*	Growth defect[c]	Shiozaki and Russell 1995
Arabidopsis	*ABI2*	Insensitive to abscisic acid	Leung et al. 1997

[a]Abbreviations: ABI, abscisic acid insensitive; ptc, phosphatase two C.
[b]Mutant phenotype is observed in *ptc1 ptp2* double mutants. *PTP2* encodes a protein tyrosine phosphatase.
[c]Mutant phenotype is observed in *ptc1Δ ptc3Δ* double mutants, and *ptc1Δ ptc2Δ ptc3Δ* triple mutants.

Table 4. PCR primers used in this work.

Name	Sequence[a]
AMC3	gatcacatcggcatcgccgtcg
AP	ggccacgcgtcgactagtacttttttttttttttt
BJM1	gsigtitwygayggicaygsigg
BJM2	acrtcccadaiiccrtcrcaigc
Cbrig3Xba	ggaatctagaaagcagtggaattctattc
Cbrig5Xho	cacctctcgagatgtccggcccaccaccac
Cele3Xba	ccgatgtcactgatctagaataactgctt
Cele5Xho	gtggttttaaaactcgagaaatacatgga
Cspe3	ggatgtcatattggatcgcg
Cspe3Nhe	cgctagcctattcattgtcatcttcttcctctagctc
Cspe5Xho	caccctctcgagatgtctgattcgctaaatc
CspefemFD	ccagatttttgtgatcggtggt
CspefemFU	acgagtcgaagaagtcggtgg
CspefemRD	gtgttcaaatcaacagcacagca
CspefemRU	tctttcgtccaatgtttctagagcc
DHA19	cgatggtcackawggtcacgag
DHA24	gaagaccaccatgaccgtgatgttgtcgtgc
DHA25	gacagacctttccgaattcaggaagttgccc
F25D1seq	ttcgatggacatgctggacaccac
F25FI	gtccacgattgcttcgctatct
F25RO	acacatgcttgatgtcacgttg
fem-2-Bgl-fast	aaatcgcgagatctttcagatagcctc
LVSL1	ggtttaattacccaagtttgag
PDPfor	gtttgccctattacatttctgtagcg
PDPrev	agtctccattctccagaacgctc
SAD2	agaaattccatcacaagccag
T3becomesT7	agctcggtaatacgactcactatagggaac
T7big	ccagtgaattgtaatacgactcactat
UAP	cuacuacuacuaggccacgcgtcgactagtac

[a]"i" represents inosine.

Table 5. *C. elegans* strains analyzed for *fem-2* and F25D1.1 polymorphisms.

Name	Geographic origin	Collected by[b]	Year collected
N2[a]	Bristol, England	L.N. Staniland	Before 1956
CB4856	Hawaii, USA	L. Hollen	1972
CB4854	Altadena, California, USA	C.D. Johnson	1973
AB1	Adelaide, South Australia, Australia	D. Riddle and A. Bird	1983
AB3	Adelaide, South Australia, Australia	D. Riddle and A. Bird	1983
KR314	Vancouver, British Columbia, Canada	F. Dill	1984
RC301	Freiburg, Germany	R. Cassada	1983
DH424	El Prieto Canyon, California, USA	L. Liao	1981
RW7000	Bergerac, France	V. Nigon	Before 1949

[a]The standard wild-type strain of *C. elegans*.

[b]For detailed information regarding the origin and history of *C. elegans* isolates see Hodgkin and Doniach (1997).

Table 6. Divergence of CB5161 and zebrafish PP2C domains.

Gene pair[a]	Protein % identity	Protein % similarity[b]	K_s [c]	K_a [d]
Cs-fem-2 and *Ce-fem-2*	68	78	1.755	0.245
Zebrafish *pp2cβ* and human PP2Cβ	73	83	1.022[e]	0.178
Zebrafish *pp2cα1* and human PP2Cα	81	87	1.035[e]	0.116
Zebrafish *pp2cα2* and human PP2Cα	80	85	1.266[e]	0.122[e]
Cs-F25D1.1 and Ce-F25D1.1	90	94	1.641	0.048
Cs-T23F11.1 and Ce-T23F11.1	95	98	0.958	0.027

[a]The GenBank accession numbers for the sequences are: *Ce-fem-2*, U29515; *Cs-fem-2*, AF177870; zebrafish *pp2cβ*, AF177869; human PP2Cβ, AJ005801; zebrafish *pp2cα1*, AF177867; human PP2Cα, AF070670; zebrafish *pp2cα2*, AF177868; Ce-F25D1.1, Z73973; Cs-F25D1.1, AF268069; Ce-T23F11.1, Z46343; and Cs-T23F11.1, AF177866.

[b]Similarity values were calculated using following similarity scheme: (I,L,V), (F,W,Y), (K,R,H), (D,E), (G,A,S), (P), (C), (T,N,Q,M).

[c]Number of synonymous substitutions per nucleotide site.

[d]Number of nonsynonymous substitutions per nucleotide site.

[e]The correction for multiple substitutions described by Kimura (1980) was not applicable, so the Jukes and Cantor (1969) correction was applied.

Table 7. Somatic phenotypes of transgenic *C. elegans fem-2(null)* animals

Construct	% with male tail (n)[a]	% with single-armed gonad (n)[a]	% without vulval tissue (n)[a]
Ce-fem-2	100(17)	100(17)	100(17)
Cb-fem-2	100(34)	100(34)	100(34)
Cs-fem-2	100(30)	100(29)	97(30)
Ce-fem-2::Cb-fem-2[b]	100(22)	100(22)	100(22)
Ce-fem-2::Cs-fem-2[c]	96(25)	96(25)	96(25)
None[d]	0(37)	0(37)	0(37)

[a]n, number of rolling XO animals examined. These animals were raised at 20°C. In the absence of a rescuing construct *fem-2(null)* m-z- XO animals are smaller than XX animals, and they have a slightly masculinized tail.

[b]The regions flanking the coding sequence in the rescuing *C. elegans* construct fused to the *C. briggsae fem-2* coding sequence.

[c]The regions flanking the coding sequence in the rescuing *C. elegans* construct fused to the CB5161 *fem-2* coding sequence.

[d]*fem-2(null)* m-z- XO progeny from non rolling hermaphrodites.

Table 8. Germ-line phenotypes of transgenic *C. elegans fem-2(b245)* animals.

Construct	% rolling XX animals laying eggs (n)[a]
Ce-fem-2	83(12)
Cb-fem-2	0(11)
Cs-fem-2	0(10)
Ce-fem-2::Cb-fem-2[b]	0(15)
Ce-fem-2::Cs-fem-2[c]	0(12)

[a]n, number of rolling XX animals examined. These animals were the F1 progeny of injected animals and were raised at 25°C.

[b]The regions flanking the coding sequence in the rescuing *C. elegans* construct fused to the *C. briggsae fem-2* coding sequence.

[c]The regions flanking the coding sequence in the rescuing *C. elegans* construct fused to the CB5161 *fem-2* coding sequence.

Table 9. Effects of RNAi against *fem-2* in *C. elegans*, *C. briggsae*, and CB5161.

Species	Injected?	% XX feminized germ line (n)[a, b]	% intersex soma (n)[a, c]	% male soma, feminized germ line (n)[a, c]	% F1 males (n)[a, c]
C. elegans	Yes	46(347)	0(120)	52(19)	37(236)
	No	0(168)	0(88)	0(12)	40(375)
C. briggsae	Yes	0(171)	8[d] (145)	48(33)	26(229)
	No	0(92)	1[d] (105)	4(55)	23(267)
CB5161	Yes	-	0(123)	0(123)	47(696)
	No	-	0(69)	0(69)	46(356)

[a]n, number of animals examined.

[b]Hermaphrodites of *C. elegans* strain N2 or *C. briggsae* strain AF16 were injected and adult progeny were examined to see whether XX animals lost the ability to produce sperm.

[c]Unmated *C. elegans him-8(e1489)* hermaphrodites or *C. briggsae mih-3(s1290)* (a high incidence of males strain) hermaphrodites were injected. CB5161 females were crossed with males prior to and following injection. Adult progeny were examined for intersex somatic characteristics, XO germ-line feminization, and sex ratio.

[d]These intersex animals had a partially-formed vulva, a wild-type male tail, and a single-armed gonad.

Table 10. Yeast two-hybrid positives obtained using FEM-2 as bait.

Construct name	Predicted gene	Encodes[a]	In frame?[b]	EST?
pDP#PS22-P5	T26C5[c]	-	-	Yes
pDP#PS22-P9	C30F8.2	Vacuolar ATPase subunit D	Yes	Yes
pDP#PS22-P12	F53F1.5	CUT 2	Yes	Yes
pDP#PS22-P20	C51G7.4	Novel protein	Yes	Yes
pDP#PS22-P21	F39D8.1	Novel protein	Yes	Yes
pDP#PS22-P23	F56A4.2	C-type lectin	Yes	Yes
pDP#PS22-P29	F56A4.2	C-type lectin	Yes	Yes

[a]A BLAST search (Altschul et al. 1997) of GenBank was performed to find sequences of known function that are similar to the predicted gene. "Novel protein" indicates that the predicted gene is believed to encode a protein that does not resemble proteins of known function. "-" indicates that the sequence is not predicted to encode a protein.

[b]Relative to the sequence coding for the GAL4 activation domain.

[c]Although the positive sequence matched a cosmid, the matching region did not correspond to a predicted gene.

Table 11. Effects of *fem-2* reporters on the fertility of *fem-2(b245)* animals.

Construct	% rolling XX animals laying eggs (n)[a]
pDP#PS1p.644	44(9)
fem-2::GFP	0(25)
fem-2::HA	27(11)
fem-2::*lacZ*	0(24)

[a]n, number of rolling XX animals examined. These animals were the F1 progeny of injected animals and were raised at 25°C.

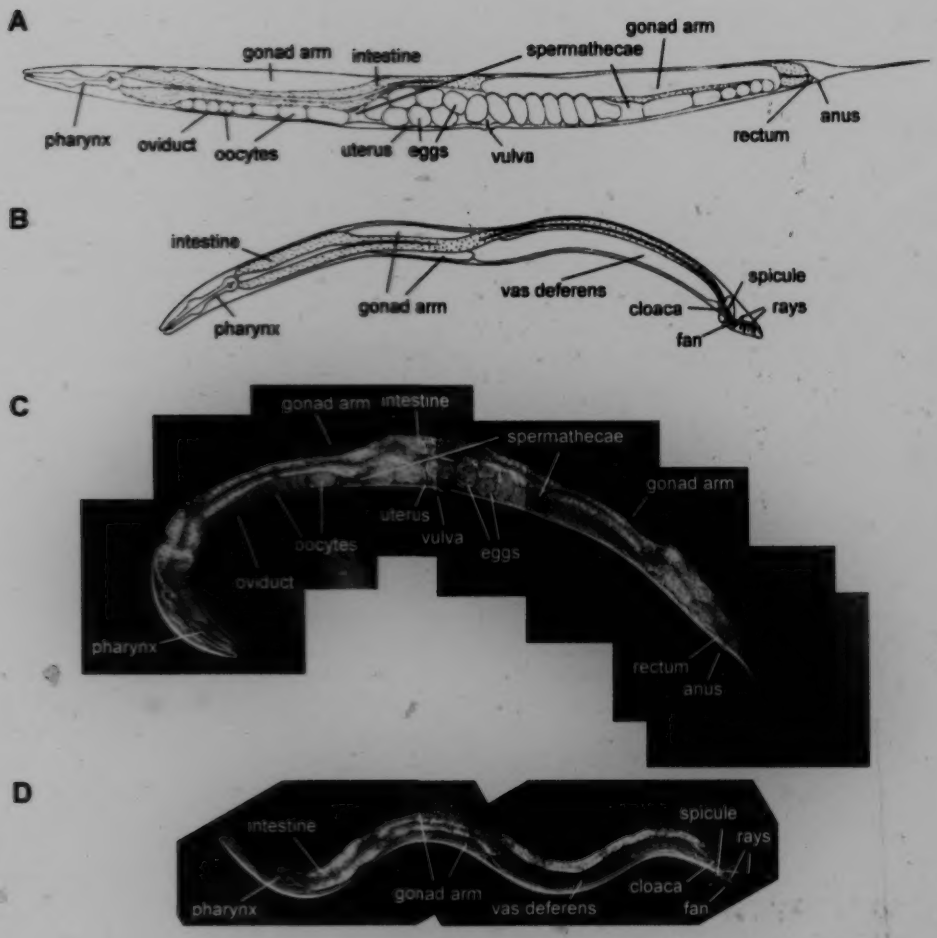

Figure 1. Overt sexual dimorphism in *C. elegans*. The drawings show the main features of the adult hermaphrodite (**A**) and male (**B**) (adapted from Sulston and Horvitz (1977)). **C**: Photograph of an adult hermaphrodite. **D**: Photograph of an adult male.

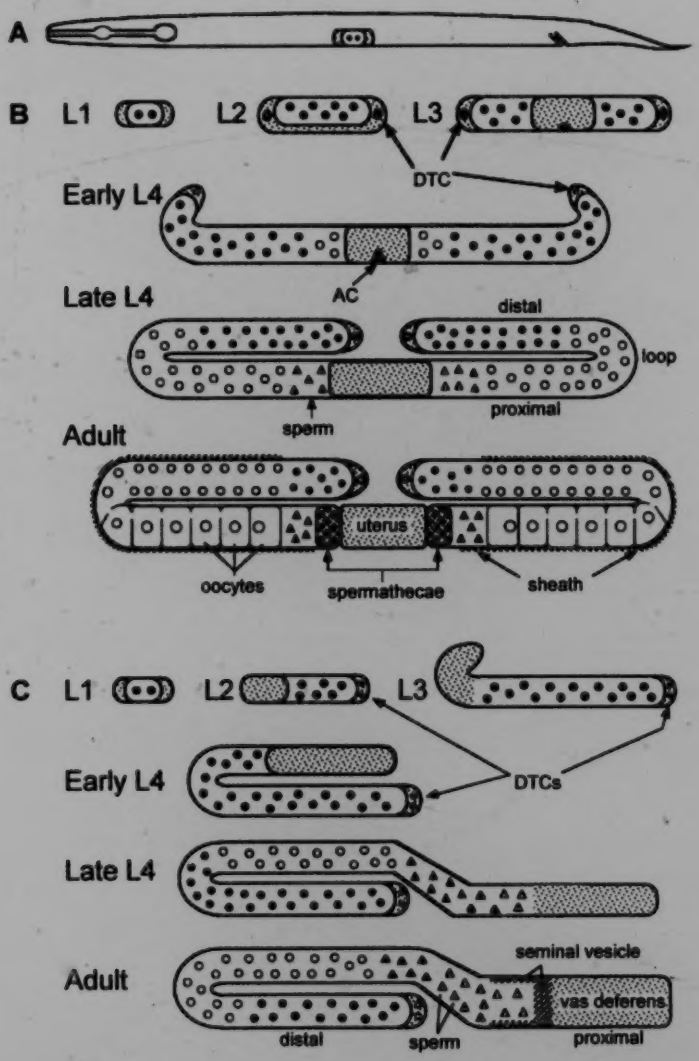

Figure 2. Development of the *C. elegans* gonad. **A**: The gonad primordium is located mid-ventrally during the first larval stage in both sexes. During subsequent larval stages there are differences in the way that the hermaphrodite (**B**) and male gonad (**C**) develop. Stippling and shading are used to highlight somatic gonad cells. Closed and open circles represent mitotic and meiotic prophase nuclei, respectively. Closed and open triangles represent primary spermatocytes and sperm, respectively. DTC, distal tip cell; AC, anchor cell. Adapted from Schedl (1997).

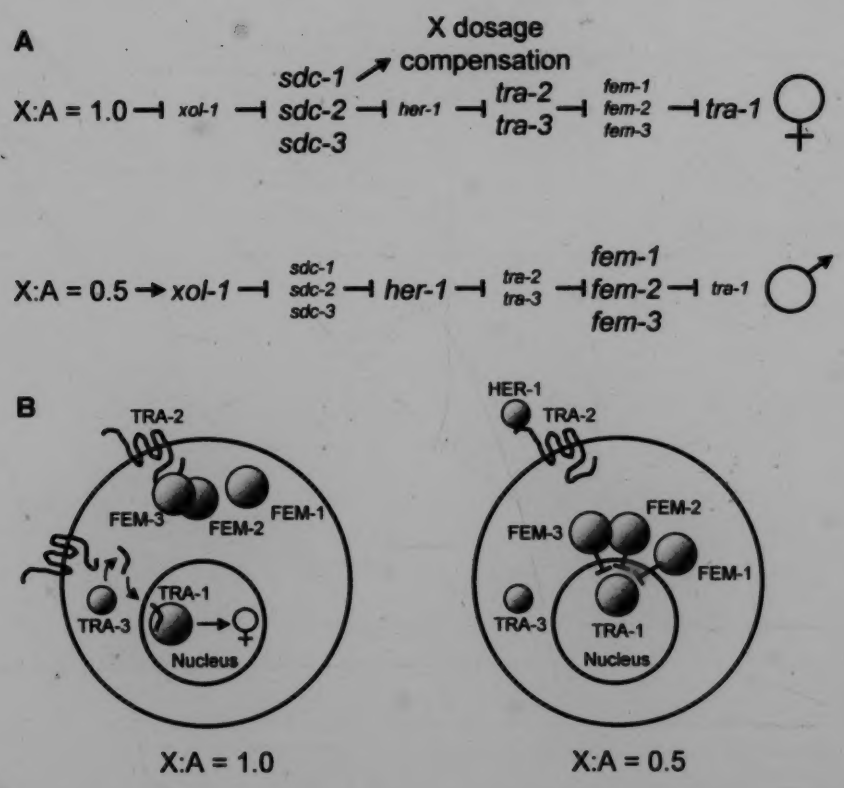

Figure 3. Model of somatic sex determination. **A**: The genetic pathway that regulates somatic sex determination in *C. elegans*. Arrows represent positive interactions, while barred lines indicate negative interactions. Gene activities that are dispensable for the given sexual fate are shown in small font. The ratio of the number of X chromosomes to the number of autosome sets (X:A ratio) is the initial signal. It establishes the sexual fate and dosage compensation modes via the *xol-1* and *sdc* genes, such that a ratio of 1.0 leads to female (hermaphrodite) development, while a ratio of 0.5 leads to male development. In XX animals *tra-2* negatively regulates the *fem* genes, allowing *tra-1* to promote female development. In XO animals *her-1* inhibits *tra-2*, allowing the *fem* genes to negatively regulate *tra-1*. In the absence of *tra-1* activity the male fate is established. **B**: The molecular model of somatic sex determination. In XX animals the TRA-2 transmembrane receptor binds to FEM-3 and interferes with the ability of the FEM proteins to inhibit the TRA-1 transcription factor. TRA-1 is then free to activate female-promoting genes and repress male-promoting genes. A portion of the cytoplasmic domain of TRA-2, released by the TRA-3 protease, may assist TRA-1. In XO animals the HER-1 secreted ligand is expressed, and it binds and inactivates TRA-2. This prevents TRA-2 from sequestering the FEMs, which are free to inactivate TRA-1.

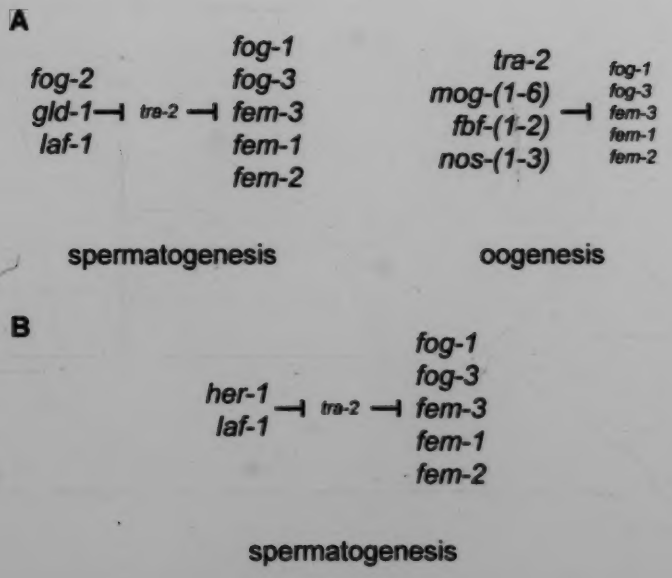

Figure 4. Genetic interactions that specify germ-line sex in *C. elegans*. Arrows represent positive interactions, while barred lines indicate negative interactions. Gene activities that are dispensable for the given sexual fate are shown in small font. In the hermaphrodite germ line (**A**), spermatogenesis requires inhibition of *tra-2* by *fog-2*, *gld-1*, and *laf-1*. The reduction in *tra-2* activity allows *fog-1*, *fog-3*, and the *fems* to function. The switch to oogenesis occurs when *tra-2* becomes active, and the *mog*, *fbf*, and *nos* genes inhibit *fem-3* translation. In the male germ line (**B**), *her-1* and *laf-1* repress the activity of *tra-2*, allowing *fog-1*, *fog-3*, and the *fems* to remain active.

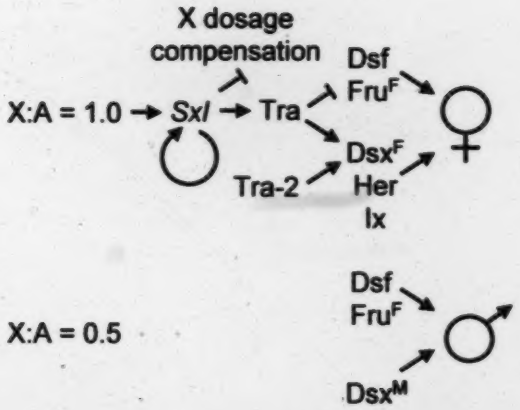

Figure 5. Somatic sex determination in *Drosophila melanogaster*. In XX animals (X:A = 1.0) *Sxl* (*Sex lethal*) is transcribed from a female-specific promoter. The resulting full-length protein contains RNA-binding motifs, and promotes its own production by modifying transcripts arising from a second non-sex-specific promoter. Sxl also blocks the translation of *msl-2* (*male-specific lethal*), and produces functional *tra* (*transformer*) transcripts. Msl-2 protein is required for dosage compensation, which is achieved by hypertranscription of X-linked genes in males. Together, Tra and Tra-2 alter the splicing of *dsx* (*doublesex*) and *fru* (*fruitless*). Tra also represses translation of the resulting *fru* mRNA, ensuring that no Fru protein is expressed. The female Dsx isoform (Dsx^F) acts with the Her (Hermaphrodite) and Ix (Intersex) proteins to promote female differentiation in the soma. The Dsf (Dissatisfaction) protein also contributes to female differentiation, specifically in the nervous system. In XY animals (X:A = 0.5) all *Sxl* transcription occurs from the non-sex-specific promoter. The resulting transcripts contain a premature stop codon, which is not removed in the absence of preexisting Sxl. As a result, functional Tra protein is not produced, and in turn the male isoforms of Dsx and Fru are expressed. Dsx^M promotes male differentiation in the soma, with the help of Fru^M and Dsf, which act in the nervous system. For more information on sex determination in *Drosophila* see reviews by Zarkower (2001), and Schutt and Nothiger (2000).

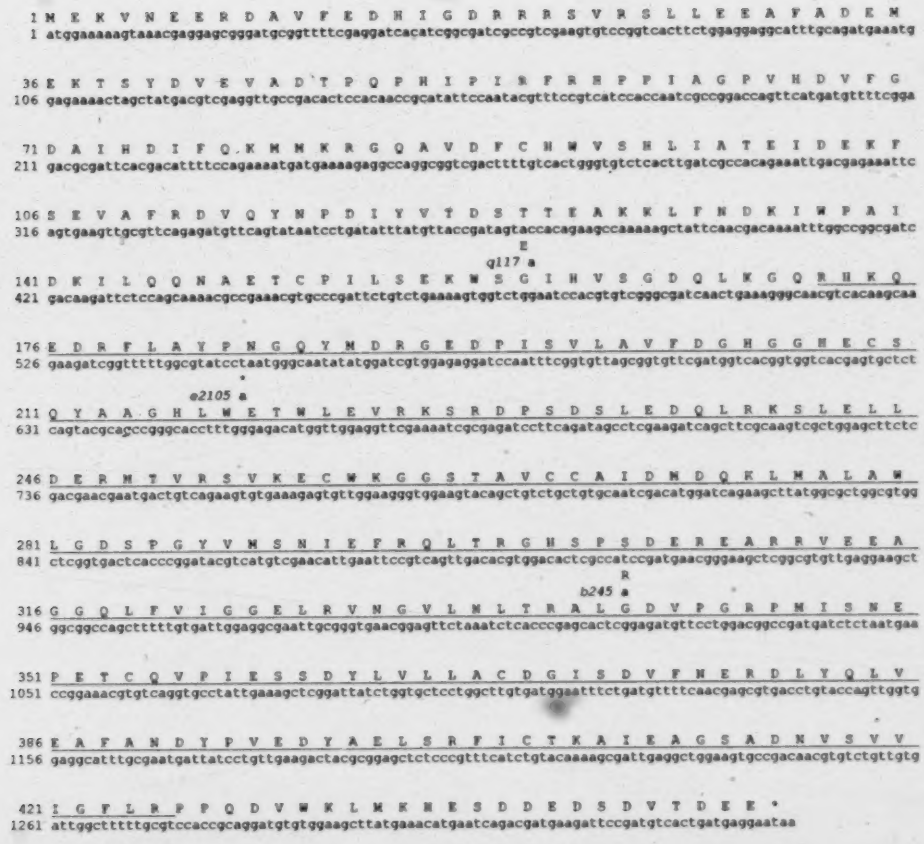

Figure 6. Mutant alleles of *fem*-2. The sequence of the *fem-2* coding region is shown in lowercase text, and its protein translation is shown in uppercase text. The residues comprising the PP2C domain described by Bork et al. (1996) are underlined. The DNA changes in the *q117*, *e2105*, and *b245* alleles are shown above the sequence, and the resulting changes in protein sequence are given. The GenBank accession number of the DNA sequence shown is U29515.

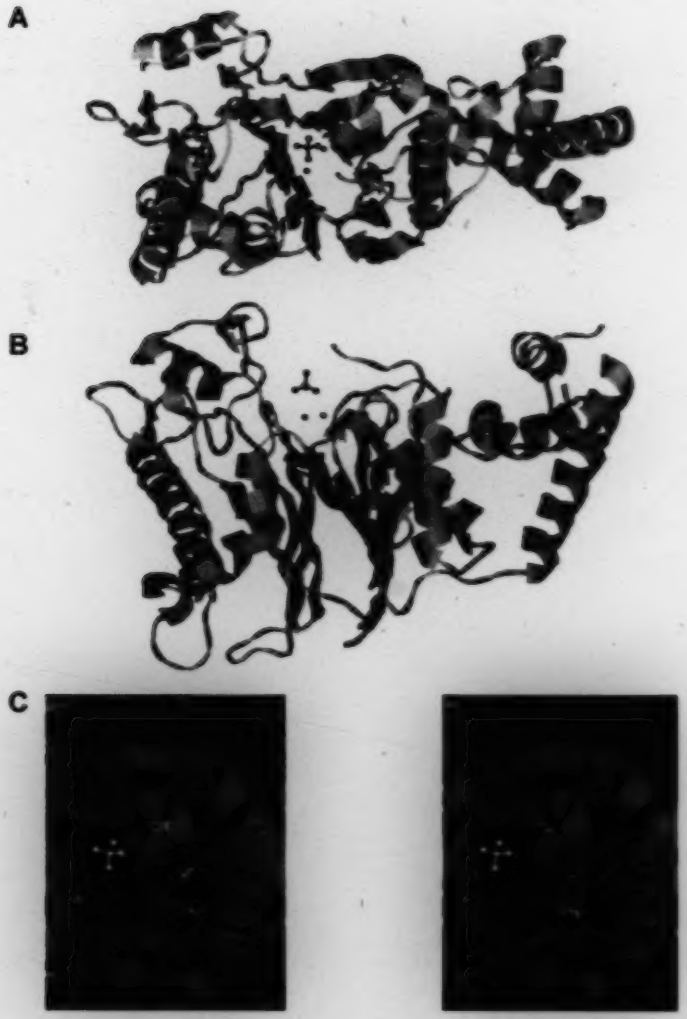

Figure 7. Three-dimensional structure of human PP2Cα. **A**: Viewed onto the catalytic site and (**B**) perpendicular to (**A**). **C**: A stereo view of the catalytic site with the metal-binding residues (E37, D38, D60, G61, D239, D282) and the phosphate-binding residue (R33) labeled. The Mn^{2+} ions (spheres) and phosphate ion are shown. The PDB ID of the file used to generate this figure is 1a6q.

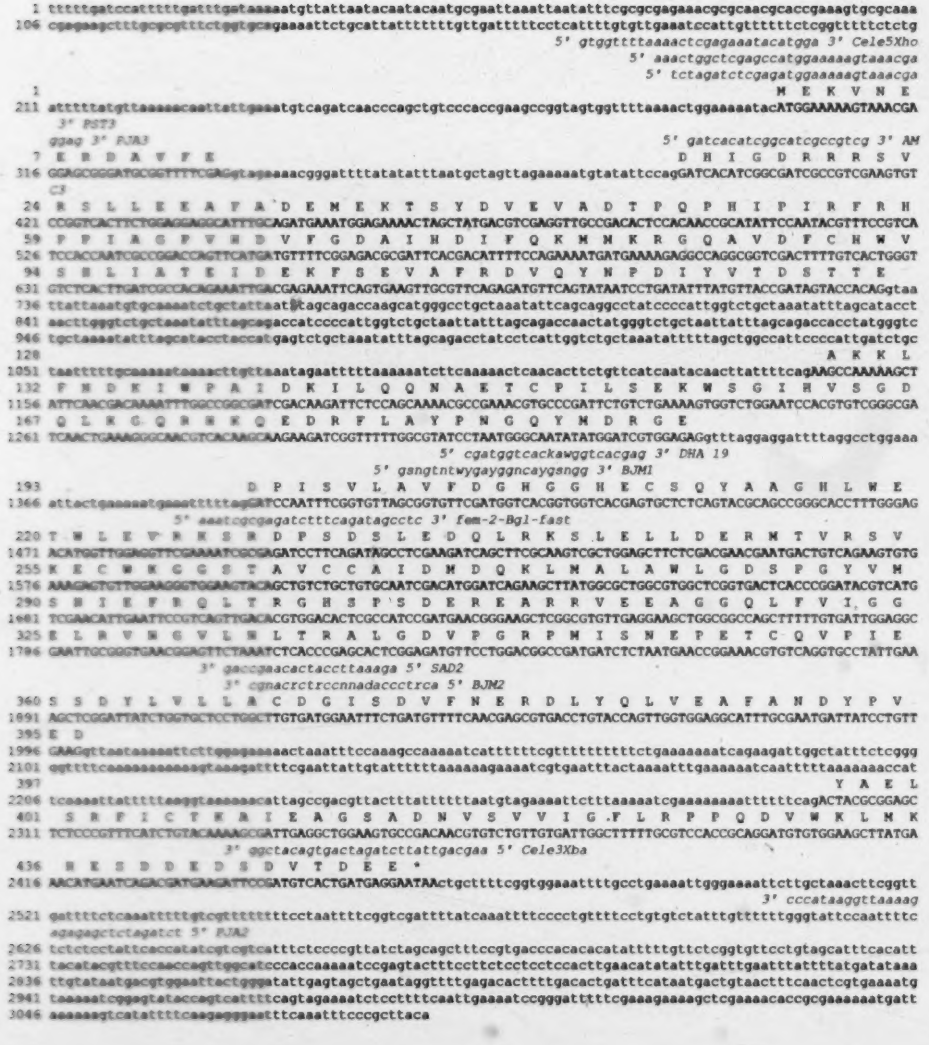

Figure 8. *C. elegans fem-2* primers. Annealing sites of primers in relation to the *C. elegans fem-2* genomic sequence are shown. The coding sequence of *fem-2* is shown in uppercase, while introns and flanking regions are shown in lowercase. The single-letter protein translation of the coding sequence is given above the coding regions. Primer names and sequences are shown in italics above putative annealing sites. The GenBank accession number of the sequence shown is U29515.

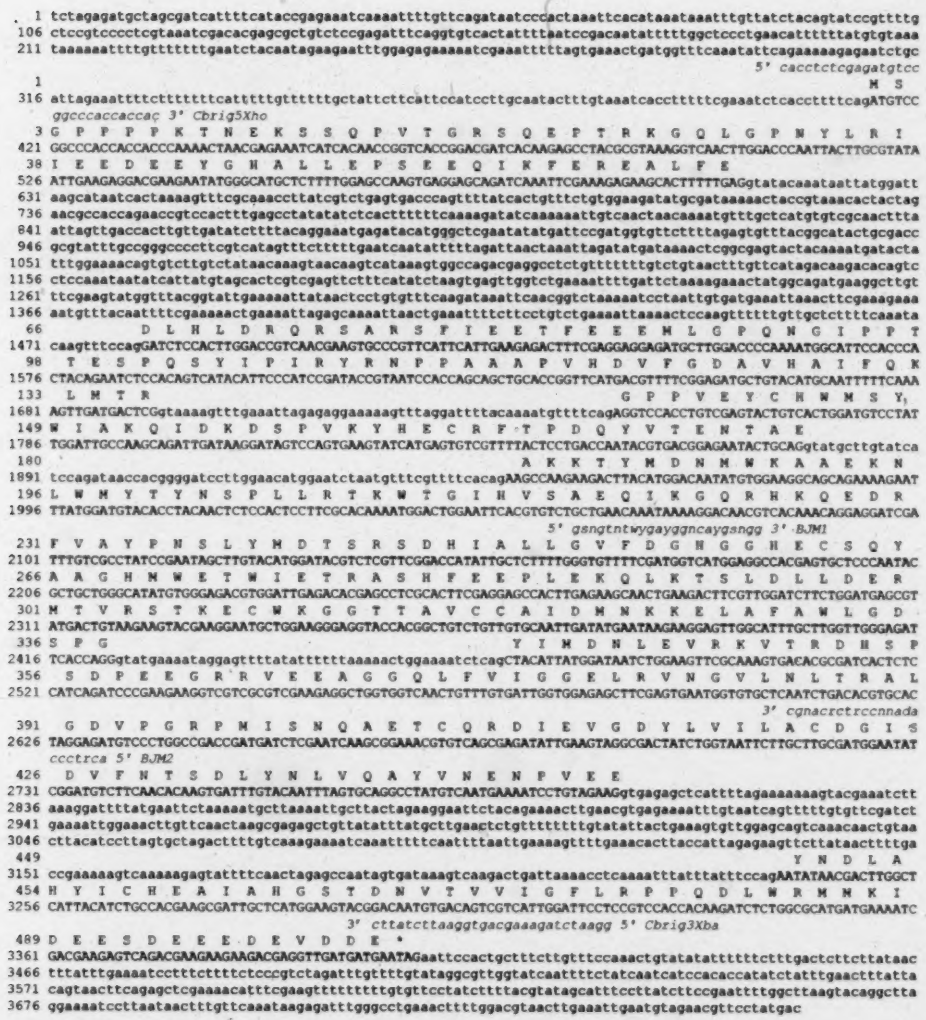

Figure 9. *C. briggsae fem-2* primers. Annealing sites of primers in relation to the *C. briggsae fem-2* genomic sequence are shown. The coding sequence of *fem-2* is shown in uppercase, while introns and flanking regions are shown in lowercase. The single-letter protein translation of the coding sequence is given above the coding regions. Primer names and sequences are shown in italics above putative annealing sites. The GenBank accession number of the sequence shown is AF054982.

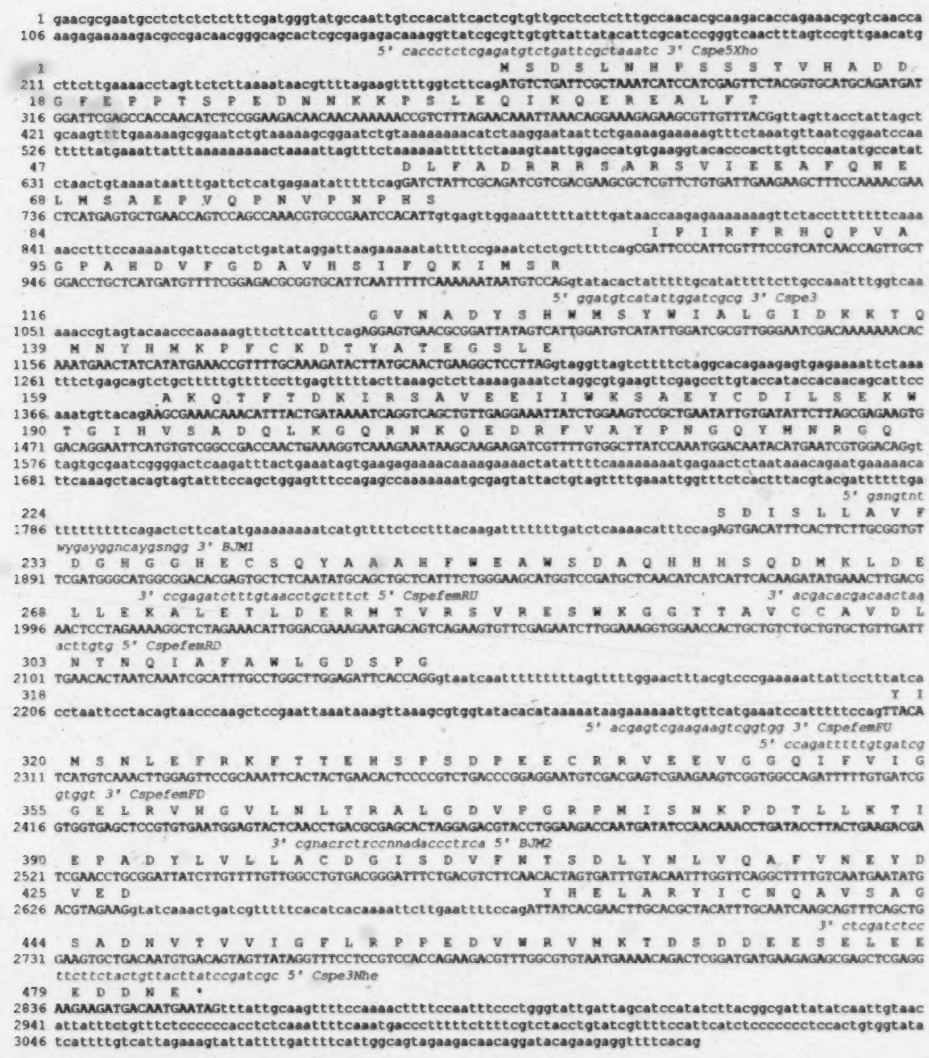

Figure 10. CB5161 *fem-2* primers. Annealing sites of primers in relation to the CB5161 *fem-2* genomic sequence are shown. The coding sequence of *fem-2* is shown in uppercase, while introns and flanking regions are shown in lowercase. The single-letter protein translation of the coding sequence is given above the coding regions. Primer names and sequences are shown in italics above putative annealing sites. The GenBank accession number of the sequence shown is AF177870.

Figure 11. Northern analysis of *Cs-fem-2* expression. A ^{32}P-labeled probe generated from a genomic fragment of *Cs-fem-2* was hybridized to 5 µg of CB5161 total RNA. The numbers indicate the relative locations of RNA size standards in kb. A single band of hybridization is seen, with an approximate size of 1.8 kb.

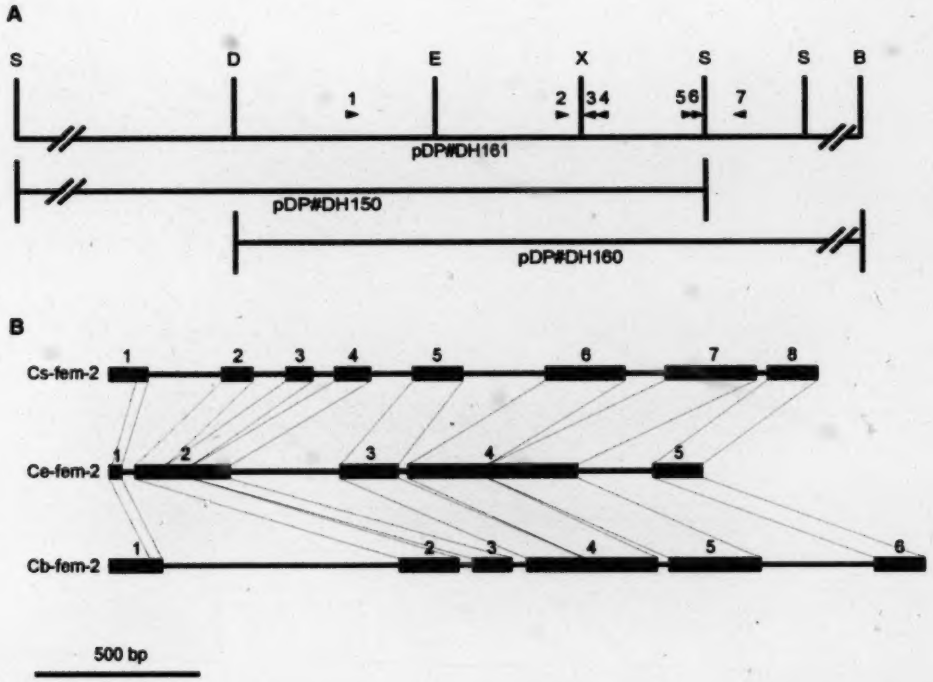

Figure 12. Genomic structure of *Cs-fem-2*. **A:** Restriction map of the genomic region containing *Cs-fem-2*. Vertical lines represent restriction sites (B, *Bam*HI; D, *Hin*dIII; E, *Eco*RI; S, *Sst*I; and X, *Xba*I), and the numbered arrows show the approximate locations of the primers used for RT-PCR and 3' RACE (1, Cspe3; 2, BJM1; 3, CspefemRU; 4, CspefemRD; 5, CspefemFU; 6, CspefemFD; and 7, BJM2). The upper horizontal line represents the DNA insert of clone pDP#DH161, while the lower lines represent clones pDP#DH150 and pDP#DH160. The hatch marks represent portions of DNA not included in this figure. The distance between the first *Sst*I site and the *Hin*dIII site is ~5.7 kbp. The distance between the last *Sst*I site and the *Bam*HI site is ~3.5 kbp. **B:** Comparison of the genomic organization of *Cs-fem-2* with *Ce-fem-2* and *Cb-fem-2*. Boxes represent exons and intervening segments represent introns. The thinnest lines show the exon relationships between orthologs. The exons and introns of *Cs-fem-2* are aligned with the restriction map shown in A.

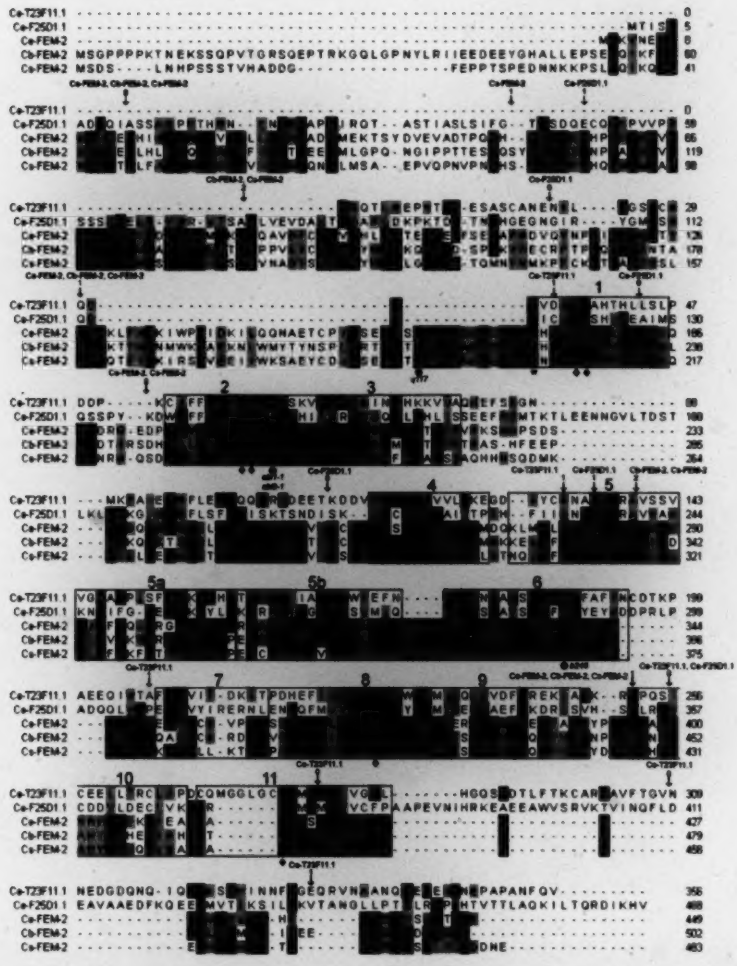

Figure 13. Alignment of FEM-2 sequences with Ce-T23F11.1 and Ce-F25D1.1. Residues that are identical among three or more sequences are shown with a black background, while similar residues are shown with a gray background using the following similarity scheme: (I,L,V), (F,W,Y), (K,R,H), (D,E), (G,A,S), (P), (C), (T,N,Q,M). The PP2C motifs described by Bork et al. (1996) are surrounded by gray boxes. Black circles mark the locations of the lesions associated with two temperature-sensitive alleles of *Ce-fem-2*, *b245* and *q117* (Pilgrim et al. 1995), and functionally mutant forms of the *Arabidopsis* PP2Cs ABI1 and ABI2 (Leung et al. 1994, 1997; Meyer et al. 1994). The black asterisk marks the residue thought to bind phosphate ions, and the black diamonds indicate residues thought to bind metal ions (Das et al. 1996). Arrows with numbers show the position and phase of introns. Each arrow is labeled with the names of the sequences that contain the intron. The GenBank accession numbers for the sequences are: Ce-T23F11.1, Z46343; Ce-F25D1.1, Z73973; Ce-FEM-2, U29515; Cb-FEM-2, AF054982; and Cs-FEM-2, AF177870.

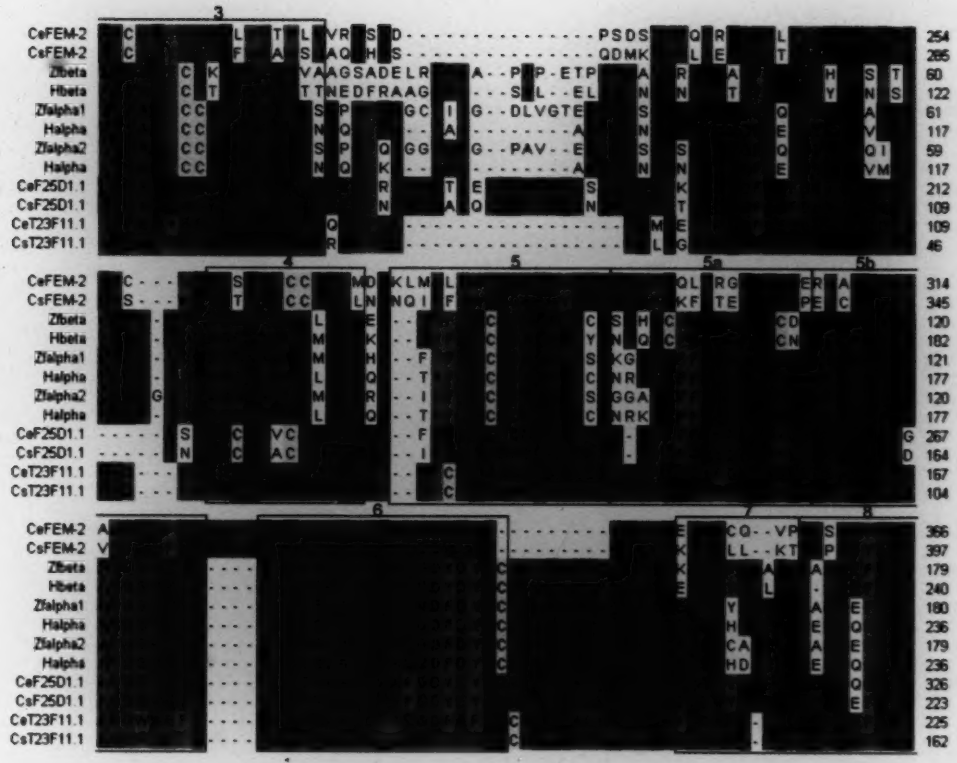

Figure 14. Alignment of the PP2C domains isolated in this study (Cs-FEM-2, Cs-T23F11.1, Cs-F25D1.1, zebrafish Pp2cβ, zebrafish Pp2cα1, zebrafish Pp2cα2) with their *C. elegans* and human orthologs. Residues that are identical or similar between ortholog pairs are shown with colored backgrounds using the following similarity scheme: (I,L,V), (F,W,Y), (K,R,H), (D,E), (G,A,S), (P), (C), (T,N,Q,M). The PP2C motifs described by Bork et al. (1996) are surrounded by gray boxes. The GenBank accession numbers for the sequences are given in Table 6.

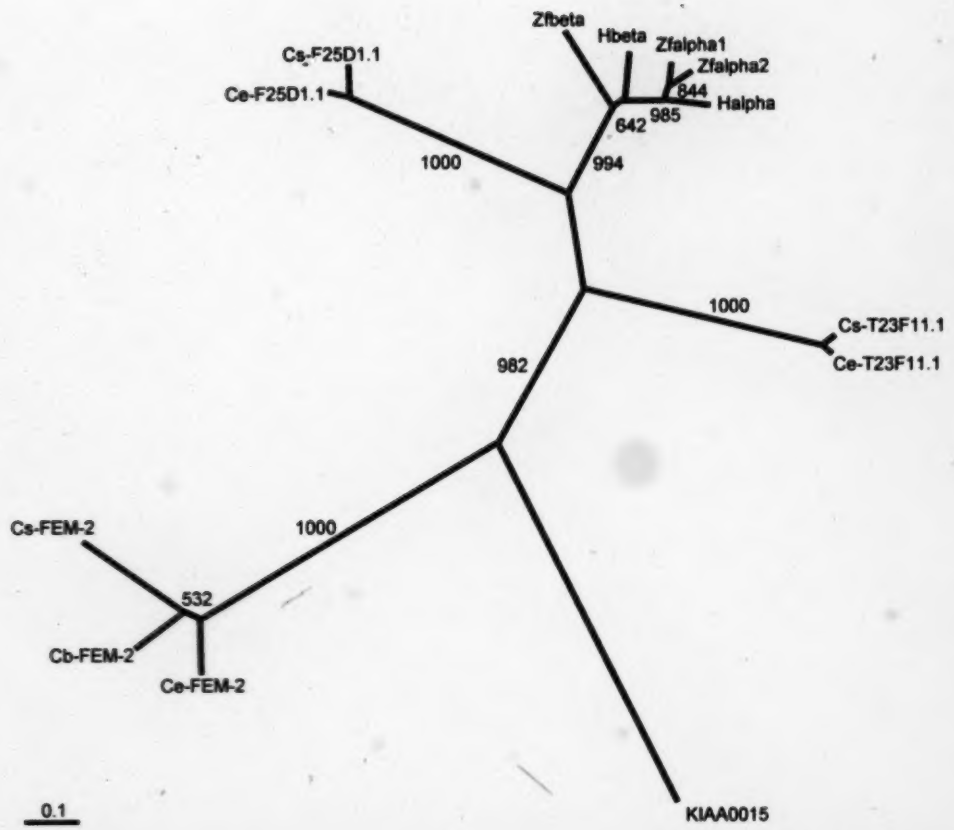

Figure 15. Phylogenetic analysis of the PP2C domains isolated in this study. Numbers at nodes indicate the bootstrap values (1000 replicates were performed). The GenBank accession numbers for the sequences are given in Table 6 and Figure 23.

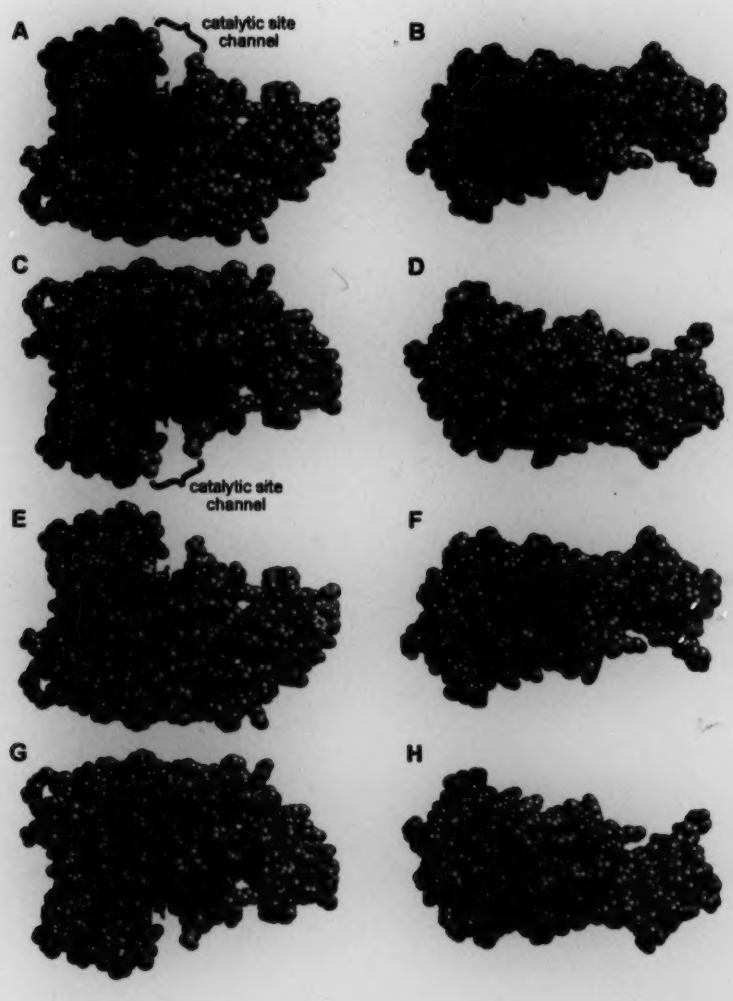

Figure 16. Conservation of FEM-2's catalytic region. **A:** View of human PP2Cα (Das et al. 1996) colored according to the conservation among the FEM-2 sequences. The phosphate and manganese ions are shown in purple. **B:** The same structure viewed onto the catalytic face. **C:** A view of the opposite side as that shown in A. **D:** A view of the opposite side as that shown in B. **E-H:** The same views as A-D but colored according to the conservation among Ce-FEM-2, Ce-F25D1.1, and Ce-T23F11.1. Residues outside of the PP2C domain are colored green, while residues within the PP2C domain that are identical or similar among all three sequences are colored dark blue and light blue, respectively. Residues that are not conserved among all three sequences are colored red. Portions of the PP2Cα sequence that align with gaps in the FEM-2 sequences are colored according to the surrounding sequence conservation (gray is used for ambiguous regions).

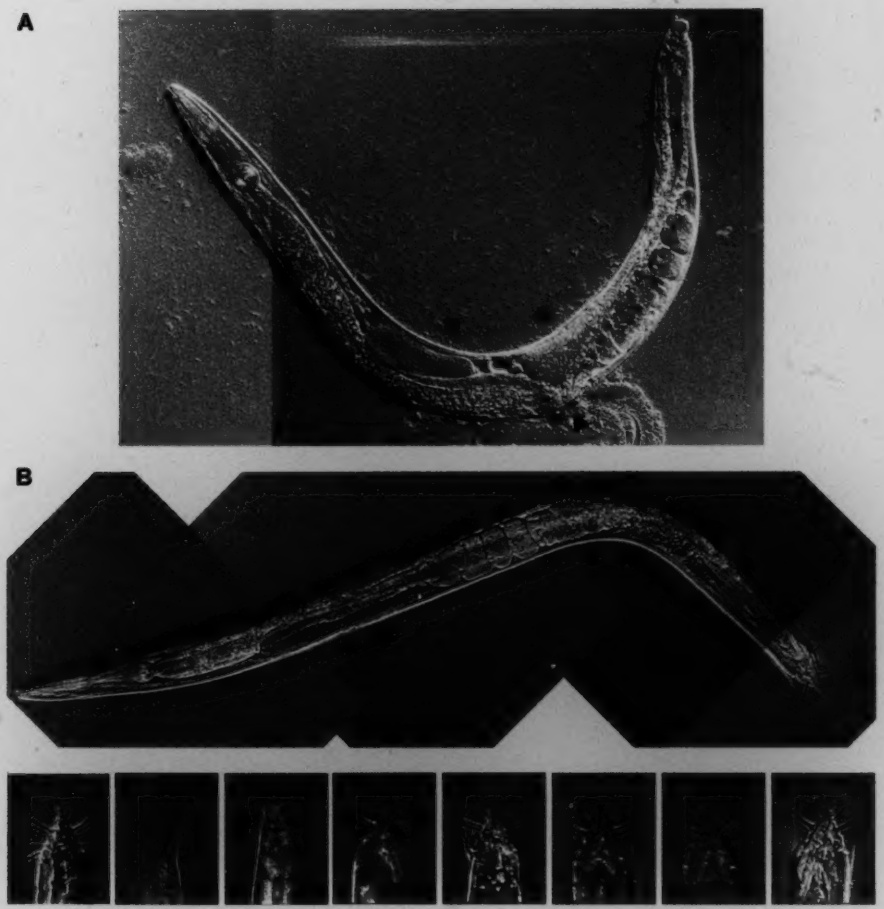

Figure 17. Rescue of somatic feminization in *C. elegans fem-2(null)* mutants. Untransformed XO m-z- *Ce-fem-2(null)* worms were compared with those carrying *fem-2* transgenic arrays. At 20°C untransformed mutant worms have a two-armed gonad, a vulva, and a slightly masculinized tail (**A**), while mutant worms carrying *fem-2* transgenes from *C.elegans*, *C. briggsae*, or CB5161 have a single-armed gonad, no vulva, and a more masculinized tail (**B**). The example shown in **B** is carrying a *Cb-fem-2* transgene fused to the *Ce-fem-2* regulatory regions. Open arrowheads point to gonad arms, and the closed arrowhead points to the vulva. **C-J:** Comparison of a wild-type male tail (**C**) and hermaphrodite tail (**D**) with mutant (**E**) and rescued (**F-J**) XO tails. The tails of males carrying a *Ce-fem-2* transgene (**F**) showed the same level of masculinization as those carrying a *fem-2* transgene from CB5161 (**G**) or *C. briggsae* (**H**), or a transgene consisting of the coding region of *Cs-fem-2* (**I**) or *Cb-fem-2* (**J**) fused to the *Ce-fem-2* flanking regions. The masculinization was incomplete, with missing rays and a reduced fan compared to the wild-type tail (**C**).

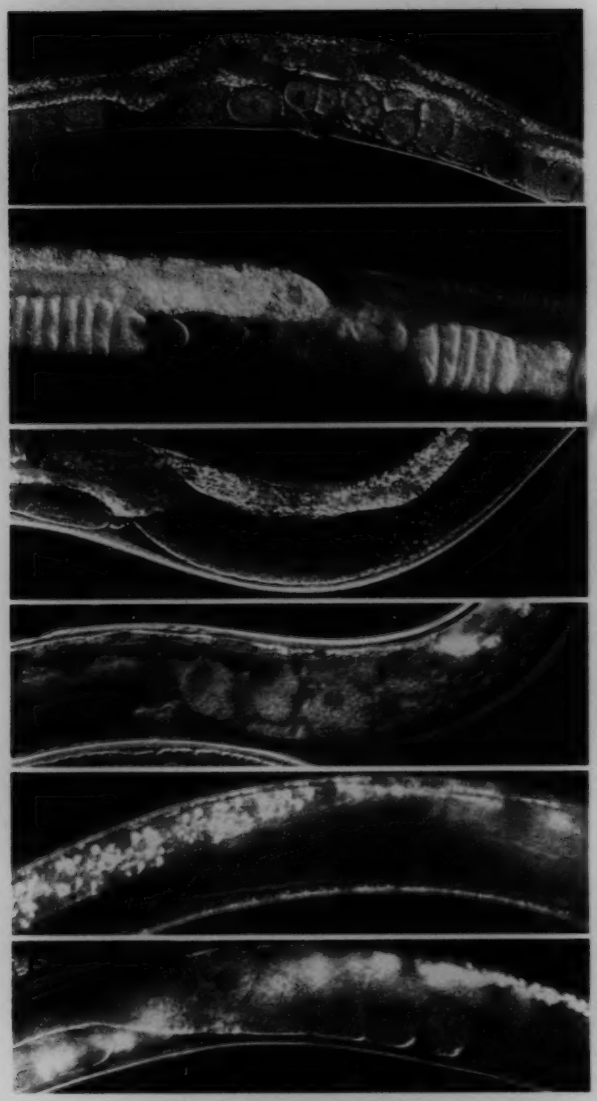

Figure 18. Results of RNAi against *fem-2*. *C. elegans* worms arising from uninjected animals appeared to be normal hermaphrodites (**A**) or males (**C**). **B**: A typical feminized *C. elegans* XX worm containing stacked oocytes and no eggs. **D**: A typical feminized *C. elegans* XO worm containing oocyte-like cells in a single-armed gonad. *C. briggsae* XX worms arising from injected animals appeared to be identical to those arising from uninjected controls (not shown). **E**: *C. briggsae* XO animals arising from uninjected animals appeared to be normal males. **F**: A typical feminized *C. briggsae* XO worm containing oocyte-like cells in a single-armed gonad.

A

```
 221 W  L  E  V  R  K  S  R  D  P  S  D  S  L  E  D  Q  L  R  K
1176 TGGTTGGAGGTTCGAAAATCGCGAGATCCTTCAGATAGCCTCGAAGATCAGCTTCGCAAG
 241 S  L  E  L  L  D  E  R  M  T  V  R  S  V  K  E  C  W  K  G
1236 TCGCTGGAGCTTCTCGACGAACGAATGACTGTCAGAAGTGTGAAAGAGTGTTGGAAGGGT
 261 G  S  T  A  V  C  C  A  I  D  M  D  Q  K  L  M  A  L
1296 GGAAGTACAGCTGTCTGCTGTGCAATCGACATGGATCAGAAGCTTATGGCGCTG
```

B

```
 246 D  E  R  M  T  V  R  S  V  K  E  C  W  K  G  G  S  T  A  V
1251 GACGAACGAATGACTGTCAGAAGTGTGAAAGAGTGTTGGAAGGGTGGAAGTACAGCTGTC
 266 C  C  A  I  D  M  D  Q  K  L  M  A  L  A  W
1342 TGCTGTGCAATCGACATGGATCAGAAGCTTATGGCGCTGGCGTGG
```

C

```
 185 T  D  S  T  L  K  L  L  E  K  G  I  K  K  G  F  L  S  F  D
1229 ACAGACAGTACACTAAAACTGCTCGAAAAGGGAATCAAGAAAGGATTCTTGAGTTTTGAT
 205 E  I  S  K  T  S  N  D  I
1289 GAAATCAGTAAAACGAGCAATGATATCgtaagttaatctgccggataacattttccactc
 214                S  K  S  G  C  T  A  V  C  A
1349 cattttatattttcagAGTAAAAGCGGATGTACAGCAGTGTGTGCC
```

Figure 19. Genomic sequence analyzed from various isolates of *C. elegans*. Coding sequences are shown in uppercase, while non coding segments are shown in lowercase. The single-letter protein translation of each coding sequence is shown. Numbers indicate the position relative to the first amino acid in the predicted protein (for the translations), or the first translated base in the gene (for the DNA sequence). **A:** The portion of the *fem-2* gene sequenced from the strains RC301, DH424, and RW7000. **B:** The portion of the *fem-2* gene sequenced from the strains HA-8, CB4854, AB1, AB3, and KR314. **C:** The portion of the F25D1.1 gene sequenced from the strains HA-8, CB4854, AB1, AB3, and KR314.

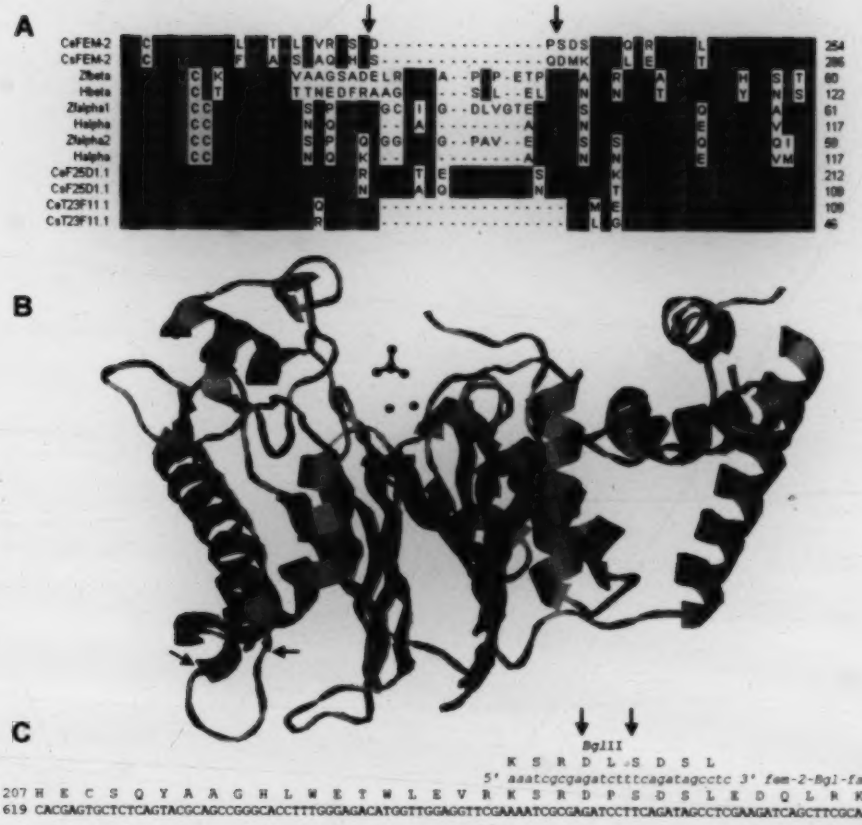

Figure 20. Site of reporter insertion in FEM-2. **A:** An alignment of CB5161 and zebrafish PP2C domains with their *C. elegans* and human orthologs. Residues that are identical or similar between orthologs are shown with colored backgrounds using the following similarity scheme: (I,L,V), (F,W,Y), (K,R,H), (D,E), (G,A,S), (P), (C), (T,N,Q,M). The arrows mark an interval that is poorly conserved. **B:** The three-dimensional structure of human PP2Cα with arrows marking the poorly-conserved interval shown in A. **C:** Partial coding sequence of *fem-2* and its protein translation. Numbers represent distance from the first coding base in the complete *fem-2* cDNA, or the first amino acid in the complete FEM-2 protein. A unique *Bgl*II site, which served as the insertion point for the reporter coding sequences, is marked above the primer used to introduce it. The interval shown in A and B is marked by arrows.

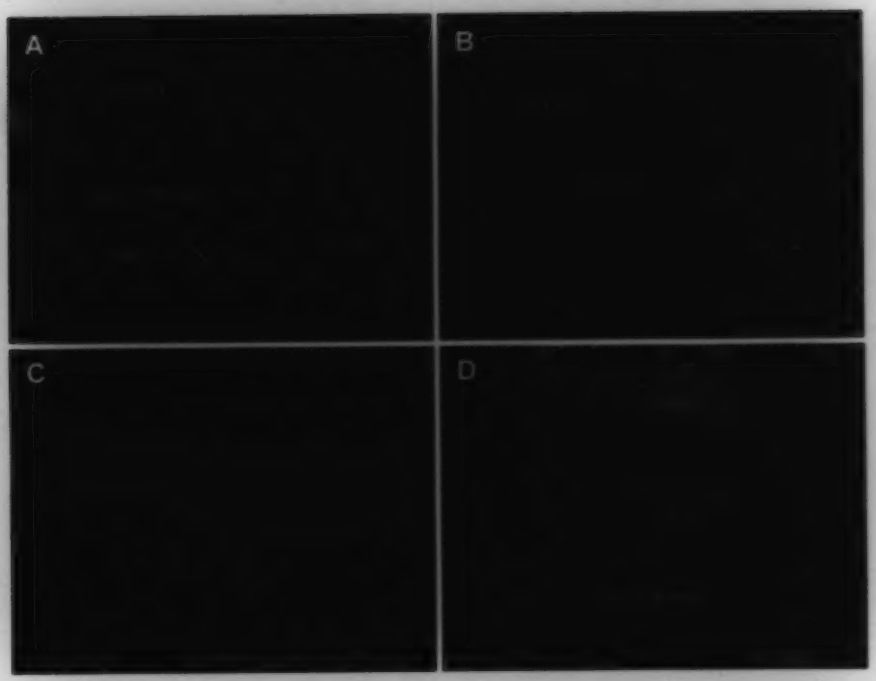

Figure 21. Expression of FEM-2::GFP fusion in N2 animals. GFP appears green, while autofluorescence appears yellow. **A:** Transgenic adult hermaphrodites typically expressed GFP in the pharynx, spermathecae, and intestine. **B**: A transgenic male expressing GFP in the pharynx and intestine. **C**: Some hermaphrodites, such as this L4 stage individual, expressed GFP in the vulva. **D**: A closer view of a hermaphrodite spermatheca.

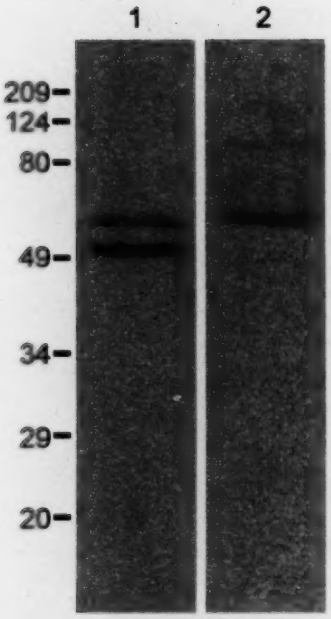

Figure 22. Expression of a FEM-2::HA fusion in *fem-2(b245)* animals. Two lanes from a western blot of proteins extracted from transgenic worms are shown. The numbers indicate the relative locations of protein size standards in kDa. Lane 1, detection performed using a polyclonal antibody raised against bacterially-expressed Ce-FEM-2. Lane 2, detection performed using a monoclonal antibody against HA.

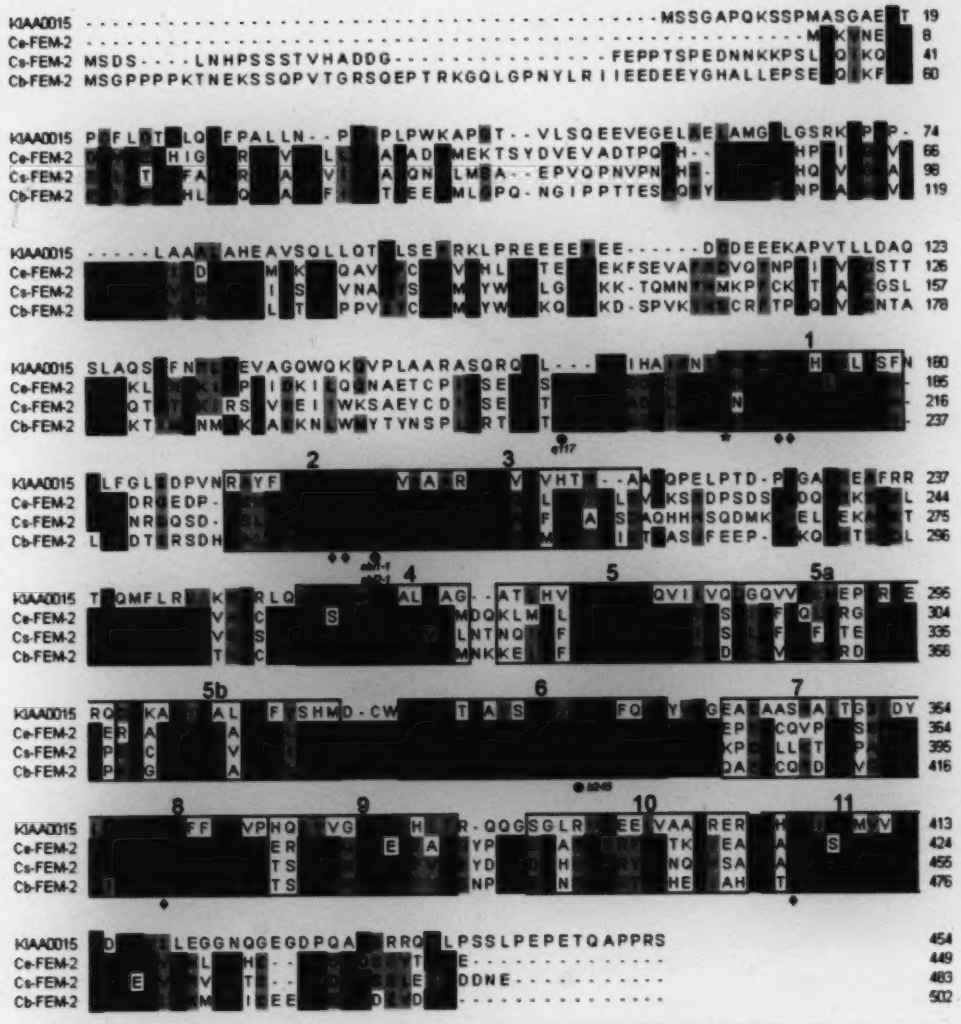

Figure 23. Alignment of FEM-2 sequences with KIAA0015. Residues that are identical among three or more sequences are shown with a black background, while similar residues are shown with a gray background using the following similarity scheme: (I,L,V), (F,W,Y), (K,R,H), (D,E), (G,A,S), (P), (C), (T,N,Q,M). The PP2C motifs described by Bork et al. (1996) are surrounded by gray boxes. Black circles mark the locations of the lesions associated with two temperature-sensitive alleles of *Ce-fem-2*, *b245* and *q117* (Pilgrim et al. 1995), and functionally mutant forms of the *Arabidopsis* PP2Cs ABI1 and ABI2 (Leung et al. 1994, 1997; Meyer et al. 1994). The black asterisk marks the residue thought to bind phosphate ions, and the black diamonds indicate residues thought to bind metal ions (Das et al. 1996). The GenBank accession number for KIAA0015 is D13640.

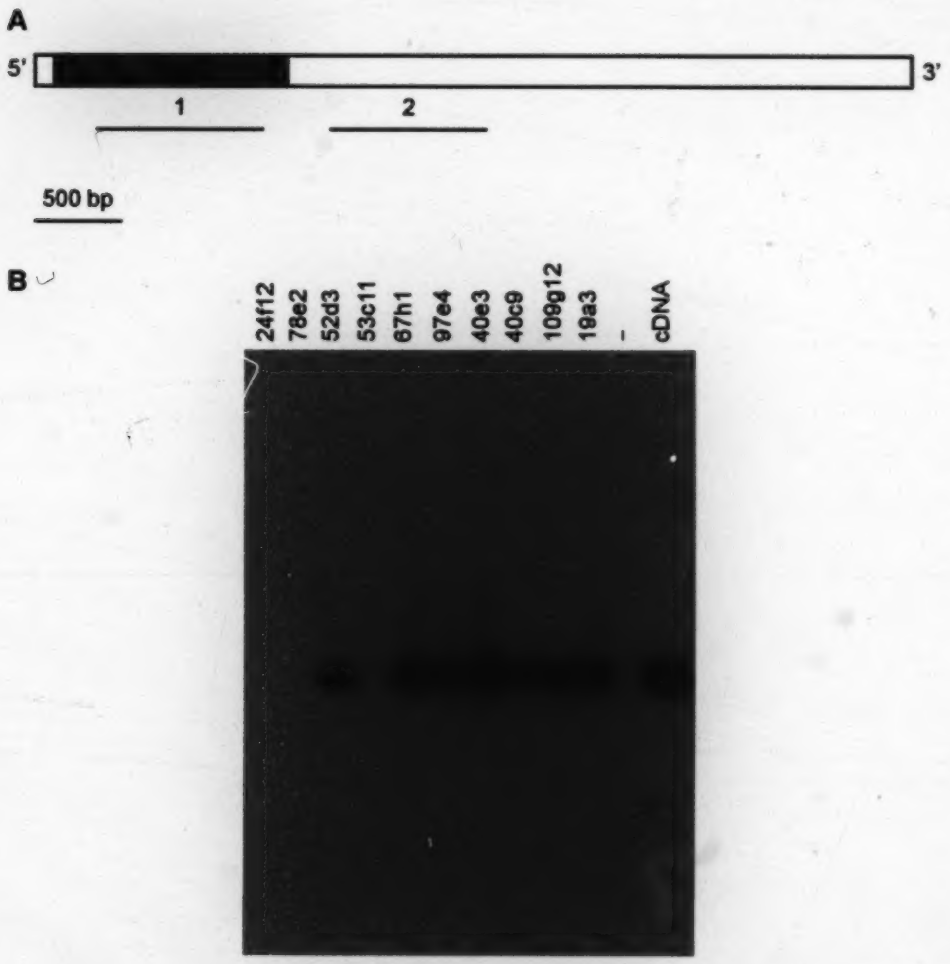

Figure 24. KIAA0015 cDNA and cosmid clones. **A**: Representation of the cDNA with the coding region indicated by a black box. The numbered lines mark the fragments used as probes (1, 970 bp PCR fragment used to isolate cosmids from library LL22NC03; 2, 922 bp *Sma*I fragment used to probe the Southern shown in B). **B**: Southern blot of *Sma*I-digested cosmid and KIAA0015 DNA probed with 922 bp *Sma*I cDNA fragment.

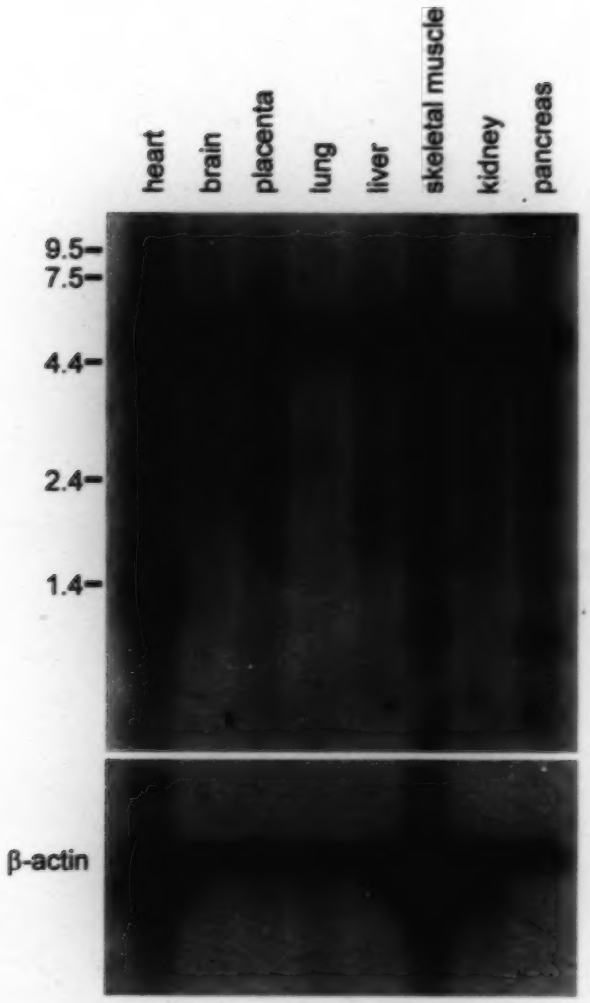

Figure 25. Northern analysis of KIAA0015 expression. A ^{32}P-labeled probe generated from the entire KIAA0015 cDNA was hybridized to poly(A)+ RNA. The numbers indicate sizes in kb. A control probing of the same blot with β-actin is shown.

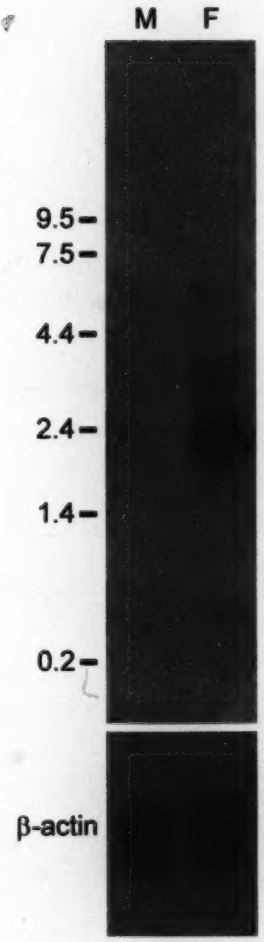

Figure 26. Expression of *pdp2* in adult zebrafish. A portion of the zebrafish *pdp2* cDNA was hybridized to total zebrafish RNA from males (M) and females (F). The numbers indicate sizes in kb. A control probing of the same blot with β-actin is shown.

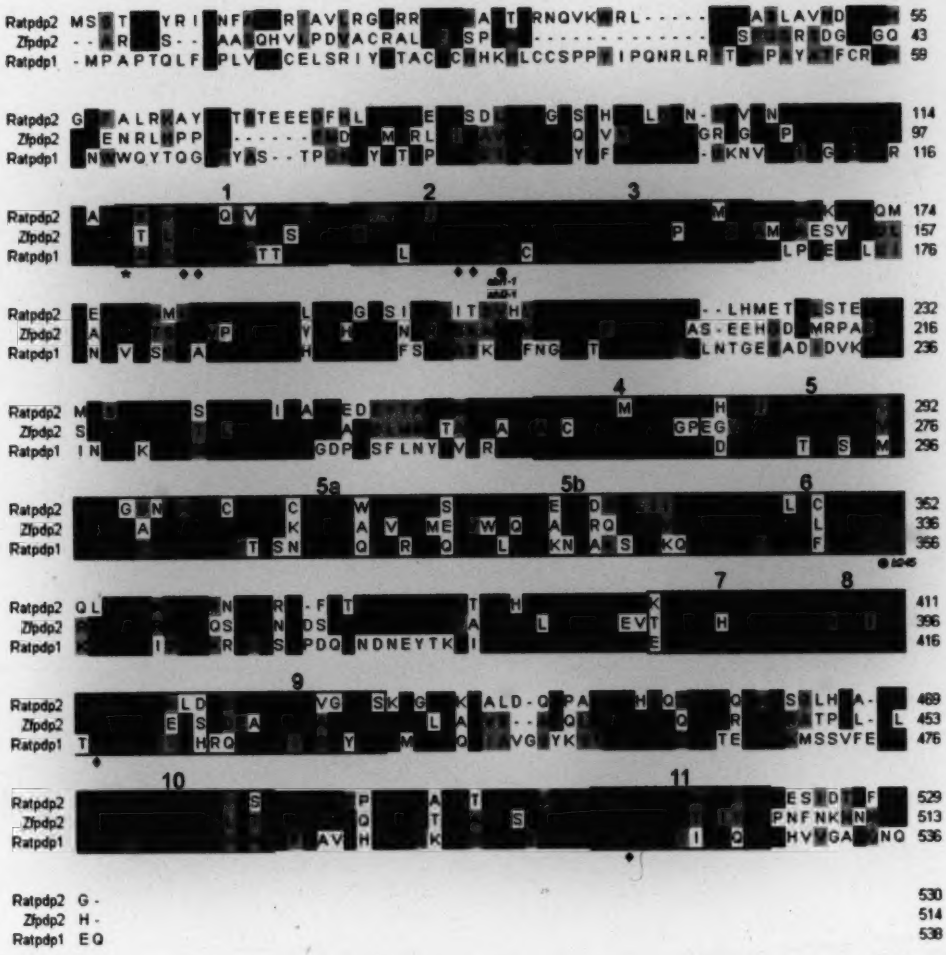

Figure 27. Alignment of zebrafish and rat PDP sequences. Residues that are identical among two or more sequences are shown with a black background, while similar residues are shown with a gray background using the following similarity scheme: (I,L,V), (F,W,Y), (K,R,H), (D,E), (G,A,S), (P), (C), (T,N,Q,M). The PP2C motifs described by Bork et al. (1996) are surrounded by gray boxes. Black circles mark the location of the lesion associated with a temperature-sensitive allele of *Ce-fem-2, b245* (Pilgrim et al. 1995), and functionally mutant forms of the *Arabidopsis* PP2Cs ABI1 and ABI2 (Leung et al. 1994, 1997; Meyer et al. 1994). The black asterisk marks the residue thought to bind phosphate ions, and the black diamonds indicate residues thought to bind metal ions (Das et al. 1996). The GenBank accession number for the sequences are: zebrafish Pdp2, AF294839; rat Pdp1, AF062740; rat Pdp2, AF062741.

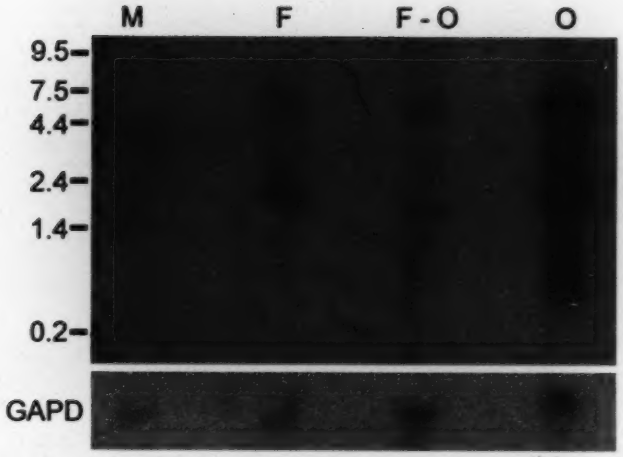

Figure 28. Zebrafish *pdp2* expression in oocytes. A portion of the zebrafish *pdp2* cDNA was hybridized to total RNA from males (M), females (F), females with oocytes removed (F – O), and oocytes (O). The numbers indicate sizes in kb. The same blot was probed with GAPD to compare loading among lanes.

Bibliography

Ahringer J, Kimble J (1991) Control of the sperm-oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3' untranslated region. Nature 349:346-348

Ahringer J, Rosenquist TA, Lawson DN, Kimble J (1992) The *Caenorhabditis elegans* sex determining gene *fem-3* is regulated post-transcriptionally. EMBO J 11:2303-2310

Akerib CC, Meyer BJ (1994) Identification of X chromosome regions in *Caenorhabditis elegans* that contain sex-determination signal elements. Genetics 138:1-21

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Barford D (1996) Molecular mechanisms of the protein serine/threonine phosphatases. Trends Biochem Sci 21:407-412

Barton MK, Kimble J (1990) *fog-1*, a regulatory gene required for specification of spermatogenesis in the germ line of *Caenorhabditis elegans*. Genetics 125:29-39

Barton MK, Schedl TB, Kimble J (1987) Gain-of-function mutations of *fem-3*, a sex-determination gene in *Caenorhabditis elegans*. Genetics 115:107-119

Barton NH, Charlesworth B (1998) Why sex and recombination? Science 281:1986-1990

Bishop JG, Dean AM, Mitchell-Olds T (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. Proc Natl Acad Sci USA 97:5322-5327

Blumenthal T, Steward K (1997) RNA processing and gene structure. In: Riddle D, Blumenthal T, Meyer B, Priess J (eds) *C. elegans* II. Cold Spring Harbor Laboratory Press, New York, pp. 117-145

Bork P, Brown NP, Hegyi H, Schultz J (1996) The protein phosphatase 2C (PP2C) superfamily: detection of bacterial homologues. Protein Sci 5:1421-1425

Brenner S, Corrochano LM (1996) Translocation events in the evolution of aminoacyl-tRNA synthetases. Proc Natl Acad Sci USA 93:8485-8489

Budarf ML, Eckman B, Michaud D, McDonald T, Gavigan S, Buetow KH, Tatsumura Y, Liu Z, Hilliard C, Driscoll D, Goldmuntz E, Meese E, Zwarthoff EC, Williams S, McDermid H, Dumanski JP, Biegel J, Bell CJ, Emanuel BS (1996) Regional localization of over 300 loci on human chromosome 22 using a somatic cell hybrid mapping panel. Genomics 35:275-288

Carmi I, Kopczynski JB, Meyer BJ (1998) The nuclear hormone receptor SEX-1 is an X-chromosome signal that determines nematode sex. Nature 396:168-173

Chan SL, Yee KS, Tan KM, Yu VC (2000) The *Caenorhabditis elegans* sex determination protein FEM-1 is a CED-3 substrate that associates with CED-4 and mediates apoptosis in mammalian cells. J Biol Chem 275:17925-17928

Chen PJ, Singal A, Kimble J, Ellis RE (2000) A novel member of the tob family of proteins controls sexual fate in *Caenorhabditis elegans* germ cells. Dev Biol 217:77-90

Cheng A, Kaldis P, Solomon MJ (2000) Dephosphorylation of human cyclin-dependent kinases by protein phosphatase type 2Cα and β2 isoforms. J Biol Chem 275:34744-34749

Cheng A, Ross KE, Kaldis P, Solomon MJ (1999) Dephosphorylation of cyclin-dependent kinases by type 2C protein phosphatases. Genes Dev 13:2946-2957

Chin-Sang ID, Spence AM (1996) *Caenorhabditis elegans* sex-determining protein FEM-2 is a protein phosphatase that promotes male development and interacts directly with FEM-3. Genes Dev 10:2314-2325

Clifford R, Lee MH, Nayak S, Ohmachi M, Giorgini F, Schedl T (2000) FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA binding protein and directs male sex determination in the *C. elegans* hermaphrodite germline. Development 127:5265-5276

Cline TW, Meyer BJ (1996) Vive la difference: males vs. females in flies and worms. Annu Rev Genet 30:637-702

Cohen P (1989) The structure and regulation of protein phosphatases. Annu Rev Biochem 58:453-508

Comeron JM (1995) A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J Mol Evol 41:1152-1159

Comeron JM (1999) K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics 15:763-764

Conradt B, Horvitz HR (1999) The TRA-1A sex determination protein of *C. elegans* regulates sexually dimorphic cell deaths by repressing the *egl-1* cell death activator gene. Cell 98:317-327

Crabtree GR, Comeau CM, Fowlkes DM, Fornace AJ Jr, Malley JD, Kant JA (1985) Evolution and structure of the fibrinogen genes. Random insertion of introns or selective loss? J Mol Biol 185:1-19

Das AK, Helps NR, Cohen PT, Barford D (1996) Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 A resolution. EMBO J 15:6798-6809

Davis TL, Meyer BJ (1997) SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome. Development 124:1019-1031

Dawes HE, Berlin DS, Lapidus DM, Nusbaum C, Davis TL, Meyer BJ (1999) Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate. Science 284:1800-1804

de Bono M, Hodgkin J (1996) Evolution of sex determination in *Caenorhabditis*: unusually high divergence of *tra-1* and its functional consequences. Genetics 144:587-595

del Peso L, Gonzalez VM, Inohara N, Ellis RE, Nunez G (2000) Disruption of the CED-9.CED-4 complex by EGL-1 is a critical step for programmed cell death in *Caenorhabditis elegans*. J Biol Chem 275:27205-27211

DeLong L, Plenebisch JD, Klein RD, Meyer BJ (1993) Feedback control of sex determination by dosage compensation revealed through *Caenorhabditis elegans sdc-3* mutations. Genetics 133:875-896

Donahue LM, Quarantillo BA, Wood WB (1987) Molecular analysis of X chromosome dosage compensation in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 84:7600-7604

Doniach T, Hodgkin J (1984) A sex-determining gene, *fem-1*, required for both male and hermaphrodite development in *Caenorhabditis elegans*. Dev Biol 106:223-235

Ellis RE, Kimble J (1995) The *fog-3* gene and regulation of cell fate in the germ line of *Caenorhabditis elegans*. Genetics 139:561-577

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806-811

Fiscella M, Zhang H, Fan S, Sakaguchi K, Shen S, Mercer WE, Vande Woude GF, O'Connor PM, Appella E (1997) Wip1, a novel human protein phosphatase that is induced in response to ionizing radiation in a p53-dependent manner. Proc Natl Acad Sci USA 94:6048-6053

Fitch DH, Bugaj-Gaweda B, Emmons SW (1995) 18S ribosomal RNA gene phylogeny for some Rhabditidae related to *Caenorhabditis*. Mol Biol Evol 12:346-358

Forgac M (1999) Structure and properties of the clathrin-coated vesicle and yeast vacuolar V-ATPases. J Bioenerg Biomembr 31:57-65

Francis R, Barton MK, Kimble J, Schedl T (1995a) *gld-1*, a tumor suppressor gene required for oocyte development in *Caenorhabditis elegans*. Genetics 139:579-606

Francis R, Maine E, Schedl T (1995b) Analysis of the multiple roles of *gld-1* in germline development: interactions with the sex determination cascade and the *glp-1* signaling pathway. Genetics 139:607-630

Gallegos M, Ahringer J, Crittenden S, Kimble J (1998) Repression by the 3' UTR of *fem-3*, a sex-determining gene, relies on a ubiquitous *mog*-dependent control in *Caenorhabditis elegans*. EMBO J 17:6337-6347

Goodwin EB, Hofstra K, Hurney CA, Mango S, Kimble J (1997) A genetic pathway for regulation of *tra-2* translation. Development 124:749-758

Goodwin EB, Okkema PG, Evans TC, Kimble J (1993) Translational regulation of *tra-2* by its 3' untranslated region controls sexual identity in *C. elegans*. Cell 75:329-339

Graham PL, Kimble J (1993) The *mog-1* gene is required for the switch from spermatogenesis to oogenesis in *Caenorhabditis elegans*. Genetics 133:919-931

Graham PL, Schedl T, Kimble J (1993) More *mog* genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of *Caenorhabditis elegans*. Dev Genet 14:471-484

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714-2723

Guthridge MA, Bellosta P, Tavoloni N, Basilico C (1997) FIN13, a novel growth factor-inducible serine-threonine phosphatase which can inhibit cell cycle progression. Mol Cell Biol 17:5485-5498

Haag ES, Kimble J (2000) Regulatory elements required for development of *Caenorhabditis elegans* hermaphrodites are conserved in the *tra-2* homologue of *C. remanei*, a male/female sister species. Genetics 155:105-116

Hall DH, Winfrey VP, Blaeuer G, Hoffman LH, Furuta T, Rose KL, Hobert O, Greenstein D (1999) Ultrastructural features of the adult hermaphrodite gonad of *Caenorhabditis elegans*: relations between the germ line and soma. Dev Biol 212:101-123

Hamilton WD, Axelrod R, Tanese R (1990) Sexual reproduction as an adaptation to resist parasites (a review). Proc Natl Acad Sci USA 87:3566-3573

Hansen D, Pilgrim D (1998) Molecular evolution of a sex determination protein: FEM-2 (PP2C) in *Caenorhabditis*. Genetics 149:1353-1362

Hill RE, Hastie ND (1987) Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature 326:96-99

Hishiya A, Ohnishi M, Tamura S, Nakamura F (1999) Protein phosphatase 2C inactivates F-actin binding of human platelet moesin. J Biol Chem 274:26705-26712

Hodgkin J (1980) More sex-determination mutants of *Caenorhabditis elegans*. Genetics 96:649-664

Hodgkin J (1986) Sex determination in the nematode *C. elegans*: analysis of *tra-3* suppressors and characterization of *fem* genes. Genetics 114:15-52

Hodgkin J (1987a) Primary sex determination in the nematode *C. elegans*. Development 101(Suppl.):5-16

Hodgkin J (1987b) A genetic analysis of the sex-determining gene, *tra-1*, in the nematode *Caenorhabditis elegans*. Genes Dev 1:731-745

Hodgkin J, Albertson DG (1995) Isolation of dominant XO-feminizing mutations in *Caenorhabditis elegans*: new regulatory *tra* alleles and an X chromosome

duplication with implications for primary sex determination. Genetics 141:527-542

Hodgkin J, Barnes TM (1991) More is not better: brood size and population growth in a self-fertilizing nematode. Proc R Soc Lond B Biol Sci 246:19-24

Hodgkin J, Doniach T (1997) Natural variation and copulatory plug formation in *Caenorhabditis elegans*. Genetics 146:149-164

Hodgkin JA, Brenner S (1977) Mutations causing transformation of sexual phenotype in the nematode *Caenorhabditis elegans*. Genetics 86:275-287

Huang B, Gudi R, Wu P, Harris RA, Hamilton J, Popov KM (1998) Isoenzymes of pyruvate dehydrogenase phosphatase. DNA-derived amino acid sequences, expression, and regulation. J Biol Chem 273:17680-17688

Hughes AL, Ota T, Nei M (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol Biol Evol 7:515-524

Hukriede NA, Joly L, Tsang M, Miles J, Tellis P, Epstein JA, Barbazuk WB, Li FN, Paw B, Postlethwait JH, Hudson TJ, Zon LI, McPherson JD, Chevrette M, Dawid IB, Johnson SL, Ekker M (1999) Radiation hybrid mapping of the zebrafish genome. Proc Natl Acad Sci USA 96:9745-9750

Hunter CP, Wood WB (1992) Evidence from mosaic analysis of the masculinizing gene *her-1* for cell interactions in *C. elegans* sex determination. Nature 355:551-555

Ishida A, Kameshita I, Fujisawa H (1998a) A novel protein phosphatase that dephosphorylates and regulates Ca2+/calmodulin-dependent protein kinase II. J Biol Chem 273:1904-1910

Ishida A, Okuno S, Kitani T, Kameshita I, Fujisawa H (1998b) Regulation of multifunctional Ca2+/calmodulin-dependent protein kinases by Ca2+/calmodulin-dependent protein kinase phosphatase. Biochem Biophys Res Commun 253:159-163

Jan E, Motzny CK, Graves LE, Goodwin EB (1999) The STAR protein, GLD-1, is a translational regulator of sexual identity in *Caenorhabditis elegans*. EMBO J 18:258-269

Jellie AM, Tate WP, Trotman CN (1996) Evolutionary history of introns in a multidomain globin gene. J Mol Evol 42:641-647

Jones AR, Schedl T (1995) Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. Genes Dev 9:1491-1504

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp. 21-132

Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Schmeits JL, Wang J, Shimizu N (1997) One-megabase sequence analysis of the human immunoglobulin λ gene locus. Genome Res 3:250-261

Kelly WG, Xu S, Montgomery MK, Fire A (1997) Distinct requirements for somatic and germline expression of a generally expressed *Caenorhabditis elegans* gene. Genetics 146:227-238

Kimble J, Edgar L, Hirsh D (1984) Specification of male development in *Caenorhabditis elegans*: The *fem* genes. Dev Biol 105:234-239

Kimble J, Sharrock WJ (1983) Tissue-specific synthesis of yolk proteins in *Caenorhabditis elegans*. Dev Biol 96:189-196

Kimble JE, White JG (1981) On the control of germ cell development in *Caenorhabditis elegans*. Dev Biol 81:208-219

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111-120

Kitani T, Ishida A, Okuno S, Takeuchi M, Kameshita I, Fujisawa H (1999) Molecular cloning of Ca^{2+}/calmodulin-dependent protein kinase phosphatase. J Biochem (Tokyo) 125:1022-1028

Klass M, Wolf N, Hirsh D (1976) Development of the male reproductive system and sexual transformation in the nematode *Caenorhabditis elegans*. Dev Biol 52:1-18

Koch R, van Luenen HG, van der Horst M, Thijssen KL, Plasterk RH (2000) Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. Genome Res 10:1690-1696

Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. Nature 336:435-440

Kraemer B, Crittenden S, Gallegos M, Moulder G, Barstead R, Kimble J, Wickens M (1999) NANOS-3 and FBF proteins physically interact to control the sperm-oocyte switch in *Caenorhabditis elegans*. Curr Biol 9:1009-1018

Kreitman M, Comeron JM (1999) Coding sequence evolution. Curr Opin Genet Dev 9:637-641

Kuwabara PE (1996a) A novel regulatory mutation in the *C. elegans* sex determination gene *tra-2* defines a candidate ligand/receptor interaction site. Development 122:2089-2098

Kuwabara PE (1996b) Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. Genetics 144:597-607

Kuwabara PE, Kimble J (1992) Molecular genetics of sex determination in *C. elegans*. Trends Genet 8:164-168

Kuwabara PE, Kimble J (1995) A predicted membrane protein, TRA-2A, directs hermaphrodite development in *Caenorhabditis elegans*. Development 121:2995-3004

Kuwabara PE, Okkema PG, Kimble J (1992) *tra-2* encodes a membrane protein and may mediate cell communication in the *Caenorhabditis elegans* sex determination pathway. Mol Biol Cell 3:461-473

Lassandro F, Sebastiano M, Zei F, Bazzicalupo P (1994) The role of dityrosine formation in the crosslinking of CUT-2, the product of a second cuticlin gene of *Caenorhabditis elegans*. Mol Biochem Parasitol 65:147-159

Lawson JE, Niu XD, Browning KS, Trong HL, Yan J, Reed LJ (1993) Molecular cloning and expression of the catalytic subunit of bovine pyruvate dehydrogenase phosphatase and sequence similarity with protein phosphatase 2C. Biochemistry 32:8987-8993

Leatherbarrow RJ, Fersht AR (1986) Protein engineering [review]. Protein Eng 1:7-16

Leung J, Bouvier-Durand M, Morris PC, Guerrier D, Chefdor F, Giraudat J (1994) *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phosphatase. Science 264:1448-1452

Leung J, Merlot S, Giraudat J (1997) The *Arabidopsis ABSCISIC ACID-INSENSITIVE2* (*ABI2*) and *ABI1* genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. Plant Cell 9:759-771

Leung-Hagesteijn C, Mahendra A, Naruszewicz I, Hannigan GE (2001) Modulation of integrin signal transduction by ILKAP, a protein phosphatase 2C associating with the integrin-linked kinase, ILK1. EMBO J 20:2160-2170

Luitjens C, Gallegos M, Kraemer B, Kimble J, Wickens M (2000) CPEB proteins control two key steps in spermatogenesis in *C. elegans*. Genes Dev 14:2596-2609

Lum DH, Kuwabara PE, Zarkower D, Spence AM (2000) Direct protein-protein interaction between the intracellular domain of TRA-2 and the transcription factor TRA-1A modulates feminizing activity in *C. elegans*. Genes Dev 14:3153-3165

Madl EM, Herman RK (1979) Polyploids and sex determination in *Caenorhabditis elegans*. Genetics 93:393-402

Maeda T, Wurgler-Murphy SM, Saito H (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature 369:242-245

Mann DJ, Campbell DG, McGowan CH, Cohen PT (1992) Mammalian protein serine/threonine phosphatase 2C: cDNA cloning and comparative analysis of amino acid sequences. Biochim Biophys Acta 1130:100-104

McAllister BF, McVean GA (2000) Neutral evolution of the sex-determining gene *transformer* in *Drosophila*. Genetics 154:1711-1720

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652-654

Mehra A, Gaudet J, Heck L, Kuwabara PE, Spence AM (1999) Negative regulation of male development in *Caenorhabditis elegans* by a protein-protein interaction between TRA-2A and FEM-3. Genes Dev 13:1453-1463

Meise M, Hilfiker-Kleiner D, Dubendorfer A, Brunner C, Nothiger R, Bopp D (1998) *Sex-lethal*, the master sex-determining gene in *Drosophila*, is not sex-specifically regulated in *Musca domestica*. Development 125:1487-1494

Mello CC, Kramer JM, Stinchcomb D, Ambros V (1991) Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J 10:3959-3970

Metz EC, Palumbi SR (1996) Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol Biol Evol 13:397-406

Meyer B, Casson LP (1986) *Caenorhabditis elegans* compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. Cell 47:871-881

Meyer K, Leube MP, Grill E (1994) A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. Science 264:1452-1455

Miller LM, Plenefisch JD, Casson LP, Meyer BJ (1988) *xol-1*: a gene that controls the male modes of both sex determination and X chromosome dosage compensation in *C. elegans*. Cell 55:167-183

Molin L, Mounsey A, Aslam S, Bauer P, Young J, James M, Sharma-Oates A, Hope IA (2000) Evolutionary conservation of redundancy between a diverged pair of forkhead transcription factor homologues. Development 127:4825-4835

Muller-Holtkamp F (1995) The *Sex-lethal* gene homologue in *Chrysomya rufifacies* is highly conserved in sequence and exon-intron organization. J Mol Evol 41:467-477

Murray MV, Kobayashi R, Krainer AR (1999) The type 2C Ser/Thr phosphatase PP2Cγ is a pre-mRNA splicing factor. Genes Dev 13:87-97

Nairn AC, Picciotto MR (1994) Calcium/calmodulin-dependent protein kinases. Semin Cancer Biol 5:295-303

Nicoll M, Akerib CC, Meyer B (1997) X-chromosome-counting mechanisms that determine nematode sex. Nature 388:200-204

Nigon V (1949) Les modalites de la reproduction et le determinisme de sexe chez quelgues nematodes libres. Ann Sci Nat Zool Biol Anim 11:1-132

Nomura N, Miyajima N, Sazuka T, Tanaka A, Kawarabayasi Y, Sato S, Nagase T, Seki N, Ishikawa K, Tabata S (1994) Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes

(KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. DNA Res 1:27-35

Nonet ML, Meyer BJ (1991) Early aspects of *Caenorhabditis elegans* sex determination and dosage compensation are regulated by a zinc-finger protein. Nature 351:65-68

Nusbaum C, Meyer BJ (1989) The *Caenorhabditis elegans* gene *sdc-2* controls sex determination and dosage compensation in XX animals. Genetics 122:579-593

Okkema PG, Kimble J (1991) Molecular analysis of *tra-2*, a sex determining gene in *C. elegans*. EMBO J. 10:171-176

O'Neil MT, Belote JM (1992) Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. Genetics 131:113-128

Perry MD, Li W, Trent C, Robertson B, Fire A, Hageman JM, Wood WB (1993) Molecular characterization of the *her-1* gene suggests a direct role in cell signaling during *Caenorhabditis elegans* sex determination. Genes Dev 7:216-228

Pilgrim D (1993) The genetic and RFLP characterization of the left end of linkage group III in *Caenorhabditis elegans*. Genome 36:712-724

Pilgrim D, McGregor A, Jackle P, Johnson T, Hansen D (1995) The *C. elegans* sex-determining gene *fem-2* encodes a putative protein phosphatase. Mol Biol Cell 6:1159-1171

Praitis V, Casey E, Collar D, Austin J (2001) Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. Genetics 157:1217-1226

Puoti A, Kimble J (1999) The *Caenorhabditis elegans* sex determination gene *mog-1* encodes a member of the DEAH-Box protein family. Mol Cell Biol 19:2189-2197

Puoti A, Kimble J (2000) The hermaphrodite sperm/oocyte switch requires the *Caenorhabditis elegans* homologs of PRP2 and PRP22. Proc Natl Acad Sci USA 97:3276-3281

Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, Zarkower D (2000) *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes Dev 14:2587-2595

Raymond CS, Shamu CE, Shen MM, Seifert KJ, Hirsch B, Hodgkin J, Zarkower D (1998) Evidence for evolutionary conservation of sex-determining genes. Nature 391:691-695

Rhind NR, Miller LM, Kopczynski JB, Meyer BJ (1995) *xol-1* acts as an early switch in the *C. elegans* male/hermaphrodite decision. Cell 80:71-82

Rudel D, Kimble J (2001) Conservation of *glp-1* regulation and function in nematodes. Genetics 157:639-654

Saccone G, Peluso I, Artiaco D, Giordano E, Bopp D, Polito LC (1998) The *Ceratitis capitata* homologue of the *Drosophila* sex-determining gene *Sex-lethal* is structurally conserved, but not sex-specifically regulated. Development 125:1495-1500

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425

Schedin P, Jonas P, Wood WB (1994) Function of the *her-1* gene is required for maintenance of male differentiation in adult tissues of *C. elegans*. Dev Genet 15:231-239

Schedl T (1997) Developmental genetics of the germ line. In: Riddle D, Blumenthal T, Meyer B, Priess J (eds) *C. elegans* II. Cold Spring Harbor Laboratory Press, New York, pp. 241-269

Schedl T, Graham PL, Barton MK, Kimble J (1989) Analysis of the role of *tra-1* in germline sex determination in the nematode *Caenorhabditis elegans*. Genetics 123:755-769

Schedl T, Kimble J (1988) *fog-2*, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. Genetics 119:43-61

Schutt C, Nothiger R (2000) Structure, function and evolution of sex-determining systems in Dipteran insects. Development 127:667-677

Shearman DC, Frommer M (1998) The *Bactrocera tryoni* homologue of the *Drosophila melanogaster* sex-determination gene *doublesex*. Insect Mol Biol 7:355-366

Shen MM, Hodgkin J (1988) *mab-3*, a gene required for sex-specific yolk protein expression and a male-specific lineage in *C. elegans*. Cell 54:1019-1031

Shiozaki K, Russell P (1994) Cellular function of protein phosphatase 2C in yeast. Cell Mol Biol Res 40:241-243

Shiozaki K, Russell P (1995) Counteractive roles of protein phosphatase 2C (PP2C) and a MAP kinase kinase homolog in the osmoregulation of fission yeast. EMBO J 14:492-502

Sievert V, Kuhn S, Traut W (1997) Expression of the sex-determining cascade genes *Sex-lethal* and *doublesex* in the phorid fly *Megaselia scalaris*. Genome 40:211-214

Skipper M, Milne CA, Hodgkin J (1999) Genetic and molecular analysis of *fox-1*, a numerator element involved in *Caenorhabditis elegans* primary sex determination. Genetics 151:617-631

Sokol SB, Kuwabara PE (2000) Proteolysis in *Caenorhabditis elegans* sex determination: cleavage of TRA-2A by TRA-3. Genes Dev 14:901-906

Spence AM, Coulson A, Hodgkin J (1990) The product of *fem-1*, a nematode sex-determining gene, contains a motif found in cell cycle control proteins and receptors for cell-cell interactions. Cell 60:981-990

Stoltzfus A, Logsdon JM Jr, Palmer JD, Doolittle WF (1997) Intron "sliding" and the diversity of intron positions. Proc Natl Acad Sci USA 94:10739-10744

Stone JM, Collinge MA, Smith RD, Horn MA, Walker JC (1994) Interaction of a protein phosphatase with an *Arabidopsis* serine-threonine receptor kinase. 266:793-795

Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102-1104

Stothard PM (2001) COMBOSA3D: combining sequence alignments with three-dimensional structures. Bioinformatics 17:198-199

Stothard P, Hansen D, Pilgrim D (2002) Evolution of the PP2C family in *Caenorhabditis*: rapid divergence of the sex-determining protein FEM-2. J Mol Evol In Press

Streit A, Li W, Robertson B, Schein J, Kamal IH, Marra M, Wood WB (1999) Homologs of the *Caenorhabditis elegans* masculinizing gene *her-1* in *C. briggsae* and the filarial parasite *Brugia malayi*. Genetics 152:1573-1584

Strovel ET, Wu D, Sussman DJ (2000) Protein phosphatase 2Cα dephosphorylates axin and activates LEF-1-dependent transcription. J Biol Chem 275:2399-2403

Struhl K (1995) Yeast transcriptional regulatory mechanisms. Annu Rev Genet 29:651-674

Sulston J, Horvitz HR (1977) Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. Dev Biol 56:110-156

Swanson WJ, Vacquier VD (1998) Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. Science 281:710-712

Takekawa M, Maeda T, Saito H (1998) Protein phosphatase 2Cα inhibits the human stress-responsive p38 and JNK MAPK pathways. EMBO J 17:4744-4752

Tamura S, Lynch KR, Larner J, Fox J, Yasui A, Kikuchi K, Suzuki Y, Tsuiki S (1989) Molecular cloning of rat type 2C (IA) protein phosphatase mRNA. Proc Natl Acad Sci USA 86:1796-1800

Tan KM, Chan SL, Tan KO, Yu VC (2001) The *C. elegans* sex-determining protein FEM-2 and its human homologue, hFEM-2, are Ca2+/calmodulin-dependent protein kinase phosphatases that promote apoptosis. J Biol Chem in press

Thomas WK, Wilson AC (1991) Mode and tempo of molecular evolution in the nematode *Caenorhabditis*: cytochrome oxidase II and calmodulin sequences. Genetics 128:269-279

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876-4882

Tong Y, Quirion R, Shen SH (1998) Cloning and characterization of a novel mammalian PP2C isozyme. J Biol Chem 273:35282-35290

Travis SM, Berger HA, Welsh MJ (1997) Protein phosphatase 2C dephosphorylates and inactivates cystic fibrosis transmembrane conductance regulator. Proc Natl Acad Sci USA 94:11055-11060

Travis SM, Welsh MJ (1997) PP2Cγ: a human protein phosphatase with a unique acidic domain. FEBS Lett 412:415-419

Trent C, Purnell B, Gavinski S, Hageman J, Chamberlin C, Wood WB (1991) Sex-specific transcriptional regulation of the *C. elegans* sex-determining gene *her-1*. Mech Dev 34:43-56

True JR, Haag ES (2001) Developmental system drift and flexibility in evolutionary trajectories. Evol Dev 3:109-119

Tucker PK, Lundrigan BL (1993) Rapid evolution of the sex determining locus in Old World mice and rats. Nature 364:715-717

Villeneuve AM, Meyer BJ (1990) The role of *sdc-1* in the sex determination and dosage compensation decisions in *Caenorhabditis elegans*. Genetics 124:91-114

Walthour CS, Schaeffer SW (1994) Molecular population genetics of sex determination genes: the *transformer* gene of *Drosophila melanogaster*. Genetics 136:1367-1372

Wang S, Kimble J (2001) The TRA-1 transcription factor binds TRA-2 to regulate sexual fates in *Caenorhabditis elegans*. EMBO J 20:1363-1372

Weis WI, Taylor ME, Drickamer K (1998) The C-type lectin superfamily in the immune system. Immunol Rev 163:19-34

Welihinda AA, Tirasophon W, Green SR, Kaufman RJ (1998) Protein serine/threonine phosphatase Ptc2p negatively regulates the unfolded-protein response by dephosphorylating Ire1p kinase. Mol Cell Biol 18:1967-1977

Wenk J, Trompeter HI, Pettrich KG, Cohen PT, Campbell DG, Mieskes G (1992) Molecular cloning and primary structure of a protein phosphatase 2C isoform. FEBS Lett 297:135-138

Whitfield LS, Lovell-Badge R, Goodfellow PN (1993) Rapid sequence evolution of the mammalian sex-determining gene *SRY*. Nature 364:713-715

Wilkins AS (1995) Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. BioEssays 17:71-77

Wright SC, Schellenberger U, Ji L, Wang H, Larrick JW (1997) Calmodulin-dependent protein kinase II mediates signal transduction in apoptosis. FASEB J 11:843-849

Yi W, Ross JM, Zarkower D (2000) *mab-3* is a direct *tra-1* target gene regulating diverse aspects of *C. elegans* male sexual development and behavior. Development 127:4469-4480

Zarkower D (2001) Establishing sexual dimorphism: conservation amidst diversity? Nat Rev Genet 2:175-185

Zarkower D, Hodgkin J (1992) Molecular analysis of the *C. elegans* sex-determining gene *tra-1*: a gene encoding two zinc finger proteins. Cell 70:237-249

Zarkower D, Hodgkin J (1993) Zinc fingers in sex determination: only one of the two *C. elegans* TRA-1 proteins binds DNA *in vitro*. Nucleic Acids Res 21:3691-3698

Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP (1997) A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. Nature 390:477-484

Zhu T, Dahan D, Evagelidis A, Zheng S, Luo J, Hanrahan JW (1999) Association of cystic fibrosis transmembrane conductance regulator and protein phosphatase 2C. J Biol Chem 274:29102-29107

Appendix 1

The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences.

JavaScript is an object-based scripting language that can be interpreted by most commonly used Web browsers, including Netscape Navigator and Internet Explorer. In conjunction with HTML form elements, JavaScript can be used to make flexible and easy to use applications that can be accessed by anyone connected to the Internet (Horton 1999). The Sequence Manipulation Suite (http://bioinformatics.org/sms/) is a collection of freely available JavaScript applications for molecular biologists. It consists of over 30 utilities for analyzing and manipulating sequence data, including the following:

Codon Plot – accepts a DNA sequence and generates a graphical plot consisting of a horizontal bar for each codon. The length of the bar is proportional to the frequency of the codon in the codon frequency table you enter. Codon frequency tables for numerous organisms are available on-line (http://www.kazusa.or.jp/codon/) (Nakamura 2000). Use Codon Plot to find portions of DNA sequence that may be poorly expressed, or to view a graphic representation of a codon usage table (by using a DNA sequence consisting of one of each codon type).

Codon Usage – accepts a DNA sequence and returns the number and frequency of each codon type. Since the program also compares the frequencies of codons that code for the same amino acid (synonymous codons), you can use it to assess whether a sequence shows a preference for particular synonymous codons.

CpG Islands - searches for CpG islands in a DNA sequence using the method of Gardiner-Garden and Frommer (1987). CpG islands are often found in the 5' regions of vertebrate genes, and hence this program can be used to highlight potential genes in genomic sequences.

DNA/Protein Pattern Find - accepts a sequence along with a set of search patterns and returns the number and positions of sites that match the patterns. The search patterns can contain "wild cards", allowing you to detect a variety of similar sequences using a single pattern.

DNA/Protein Stats – returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each residue, and for certain groups of residues, allowing you to quickly compare the results obtained for different sequences.

Filter DNA/Protein - removes non-protein or non-DNA characters from text. Use this program when you wish to remove digits and blank spaces from a sequence to

make it suitable for other applications. All the programs in the Sequence Manipulation Suite automatically filter the sequences you enter before proceeding.

GenBank Feature Extractor - accepts a GenBank file as input and reads the sequence feature information described in the feature table, according to the rules outlined in the GenBank release notes (ftp://ncbi.nlm.nih.gov/genbank/gbrel.txt). The program concatenates the relevant sequence segments and returns each sequence feature in FASTA format. GenBank Feature Extractor is particularly helpful when you wish to derive the sequence of a cDNA from a genomic sequence that contains many introns.

GenBank to FASTA - accepts a GenBank file as input and returns the entire DNA sequence in FASTA format. Use this program when you wish to quickly remove all of the non-DNA sequence information from a GenBank file.

GenBank Trans Extractor - accepts a GenBank file as input and returns each of the protein translations described in the file in FASTA format. GenBank Trans Extractor should be used when you are more interested in the predicted protein translations of a DNA sequence than the DNA sequence itself.

Group DNA/Protein - adjusts the spacing of DNA or protein sequences and adds numbering. You can specify the group size (the number of residues per group), as well as the number of residues per line. The output of this program can serve as a

convenient reference for use with some of the other programs in the Sequence Manipulation Suite (such as DNA/Protein Pattern Find), since the numbering and spacing allows you to quickly locate specific residues.

Ident and Sim – accepts a pair of aligned sequences (in FASTA format) and calculates their identity and similarity. Identity and similarity values are often used to assess whether or not two sequences share a common ancestor or function. Since many alignment programs do not calculate these values, you may find Ident and Sim quite useful when making pairwise comparisons.

Multiple Align Show - accepts a group of aligned sequences (in FASTA format) and formats the alignment to your specifications. You can specify the number of residues per line, and whether to use colored text or colored backgrounds to highlight matching residues. You can also set a consensus level, which specifies the percentage of residues that need to be identical in a column of the alignment for highlighting to be added. Use Multiple Align Show to enhance the output of sequence alignment programs.

ORF Finder - searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF Finder to search newly sequenced DNA for potential protein encoding segments.

Primer Show - accepts a DNA sequence along with a set of primer sequences and returns a textual map showing the annealing positions of the primers. The translation of the DNA sequence can be shown in the reading frames you specify. You can also choose the number of bases per line of the map, and whether to show the DNA in its single-stranded or double-stranded form. The primer sequences you enter can contain "wild card" bases, a feature that allows Primer Show to handle degenerate primers. Use this program to produce a useful reference figure, particularly when you have designed a large number of primers for a particular template. A more advanced version of this program, called Sequence Extractor, is now available (http://bioinformatics.org/seqext/).

Protein Molecular Weight - accepts a protein sequence and calculates the molecular weight. You can append copies of commonly used epitopes and fusion proteins using the supplied list. Use Protein Molecular Weight when you wish to predict the location of a protein of interest on a gel in relation to a set of protein standards.

Random DNA/Protein Sequence - generates a random sequence of the length you specify. Random sequences can be used to evaluate the significance of sequence analysis results.

Rest and Trans Map - accepts a DNA sequence and returns a textual map showing the positions of restriction endonuclease cut sites. The translation of the DNA sequence is also given, in the reading frame you specify. You can choose the number

of bases per line of the map, and whether to show the DNA in its single-stranded or double-stranded form. Use the output of this program as a reference when planning cloning strategies, particularly when you need to consider the reading frames of the molecules you are ligating.

Restriction Summary - accepts a DNA sequence and returns the number and positions of restriction endonuclease cut sites. Use this program if you wish to quickly determine whether or not an enzyme cuts a particular segment of DNA, and to produce a table to complement the output of Rest and Trans Map.

Reverse Complement - converts a DNA sequence into its reverse, complement, or reverse-complement counterpart. You may wish to work with the reverse-complement of a sequence if it contains an ORF on the reverse strand.

Reverse Translate – accepts a protein sequence and uses a codon usage table to generate a graph that can be used to find regions of minimal degeneracy at the nucleotide level. Use Reverse Translate to design PCR primers for amplifying genes that encode similar proteins.

Shuffle DNA/Protein - randomly shuffles a sequence. Shuffled sequences can be used to evaluate the significance of sequence analysis results, particularly when sequence composition is an important consideration.

Simple Plot - calculates DNA sequence composition using a sliding window. The results are returned as a set of x-values and y-values. The x-value is the position of the first base in the window of bases used in the calculation, and the y-value is the result of the calculation. Use Simple Plot when you are looking for segments of a DNA sequence with a particular base composition, rather than a specific sequence of bases.

TestCode – accepts a DNA sequence and calculates the TestCode value as described by Fickett (Fickett 1982). The reading frame of the DNA sequence does not need to be considered when performing this analysis. Use TestCode to predict whether or not a sequence encodes a protein.

In addition to the features described above, the interface to each application displays (where applicable) the genetic code, restriction enzyme set, and codon usage table it uses so that you can make changes before performing an operation. Each program also comes with a complete set of default inputs to help illustrate the desired data formats.

To access a program in the Sequence Manipulation Suite simply point your Web browser to (http://bioinformatics.org/sms/) and click on the name of the program you wish to use. The time needed for the applications to complete their tasks is dependent on how much input you give them. Try short sequences first to gauge how they perform on your computer.

Each program in the Sequence Manipulation Suite writes its results to the same output window, which appears when the first analysis is performed. Whenever you perform an additional analysis the results are simply appended to the existing output data. By scrolling up and down in the window you can easily compare the results of a particular operation performed on a series of sequences, or you can compare the results of many different operations performed on a single sequence. You can also copy a sequence segment of interest from the output window, and paste it back into a Sequence Manipulation Suite program for further analysis or formatting.

To save or print the contents of the output window, copy the contents and paste them into your favorite text editor. If the spacing of the text looks incorrect after it has been pasted, switch to a fixed-width font such as courier. The color and font of the text shown in the output window may not be conserved after copying and pasting, depending on which web browser and text editor you are using.

The Sequence Manipulation Suite can be used on-line, or a copy of the programs can be downloaded and saved on your own computer for use off-line (http://bioinformatics.org/sms/download.html). One advantage of using a local copy of the programs is that their default settings can be modified and saved to suit your preferences. For example, you might want to replace the codon usage tables with a table generated for the organism you study. You might also want to replace the restriction enzyme sets with a set consisting of the enzymes you have access to in your lab. Anyone with HTML experience is encouraged to make these changes. Readers familiar with JavaScript may also want to build new programs by combining and modifying existing portions of the Sequence Manipulation Suite.

References:

Fickett JW (1982) Recognition of protein coding regions in DNA sequences. Nucleic Acids Res 10:5303-5318

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196:261-282

Horton RM (1999) Biological sequence analysis using regular expressions. BioTechniques 27:76-78

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucleic Acids Res 28:292

Appendix 2

COMBOSA3D: combining sequence alignments with three-dimensional structures

If you flip through any current journal that deals with molecular biology you are likely to see several sequence alignments. These figures are popular because they illustrate potential evolutionary and functional relationships among sequences in an easy to interpret format. In cases where one of the sequences has a known three-dimensional structure it can be informative to compare the alignment with the solved structure, to better understand how the local environment of the residues relates to conservation. This type of comparison often involves preparing a sequence alignment figure in which the secondary structure characteristics of the solved sequence are added manually as drawings. Another method involves adjusting the font of the solved sequences in an alignment figure to reflect particular characteristics extracted from the three-dimensional structure (Mizuguchi et al. 1998). Here I describe COMBOSA3D (Coloring Of Molecules Based On Sequence Alignment), a program that combines sequence alignment information with three-dimensional structures in an alternative and informative way. Instead of adjusting the appearance of a sequence alignment figure to convey structural information, COMBOSA3D adjusts a three-dimensional structure to convey alignment information. The resulting annotated structure can be used to address questions regarding the functional significance of evolutionary and mutational differences in an intuitive way. The program can be tested at http://bioinformatics.org/combosa3d/, and the source code is freely available.

COMBOSA3D is primarily mouse and menu driven and it runs in popular Web browsers, making it accessible to a wide range of researchers. The interface to the program consists of a single window divided into four resizable viewing areas, or panes. One of the panes uses the Chime molecular structure viewing plug-in (available from http://www.mdli.com/download/) to display three-dimensional structures, while the others are used primarily for data entry and structure manipulation. Operation of the program is divided into three basic steps, beginning with the loading of a protein structure of interest. Protein structures are obtained remotely from the Protein Data Bank (Berman et al. 2000) using a Protein Data Bank ID, which is entered in the Molecule Control pane. To manipulate the orientation and display of the loaded structure a set of buttons and menus is available, along with a text area that accepts Chime and RasMol (Sayle and Milner-White 1995) scripts. Next, the pre-aligned set of sequences that is to be compared with the loaded three-dimensional structure is entered into the Alignment text area, located in the Sequence pane. Finally, the loaded molecule's sequence (which may or may not be included in the Alignment text area) is pasted in aligned form into the Solved Sequence text area. The program then evaluates each column in the alignment entered into the Alignment text area, to determine the level of sequence similarity, and a color is assigned to the aligned residue in the solved sequence, depending on the level of sequence similarity detected. After all the columns have been evaluated the resulting color scheme is applied to the loaded structure (Figure 1A). A formatted textual alignment colored using the same color scheme is also generated to serve as a reference (Figure 1B). Pointing to a residue in the solved sequence in the reference alignment displays the

residue's position in the Web browser's status bar. Alignment segments of interest can be labeled on the three-dimensional structure by entering residue positions or ranges into the Molecule Control pane. Alternatively, specific residues in the three-dimensional structure can be identified in the status bar by clicking on the structure.

A variety of options are available to adjust the way that COMBOSA3D calculates color schemes. The percentage threshold for the coloring of indels (amino acids in the solved sequence that align with gaps in the other sequences), identical, and similar residues can be set, and the user can define which groups of residues qualify as similar. The colors used to represent indels, identical, similar, and divergent residues can also be specified. A range of gray colors is included in the color options so that informative black and white figures can be generated. The molecule and the reference alignment can be re-colored using new settings at any time by clicking the Color Molecule and Show Alignment buttons, respectively. Once the desired view of the colored molecule is obtained, a picture of the molecule can be copied to the Clipboard (using the Clipboard button) for pasting into a graphics program, and the reference alignment can be saved as an HTML file or printed from within the Web browser.

The author would like to thank Eric Martz (University of Massachusetts) for helpful comments relating to COMBOSA3D.

References:

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235-242

Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 265:346-355

Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP (1998) JOY: protein sequence-structure representation and analysis. Bioinformatics 14:617-623

Sayle RA, Milner-White EJ (1995) RASMOL: biomolecular graphics for all. Trends Biochem Sci 20:374

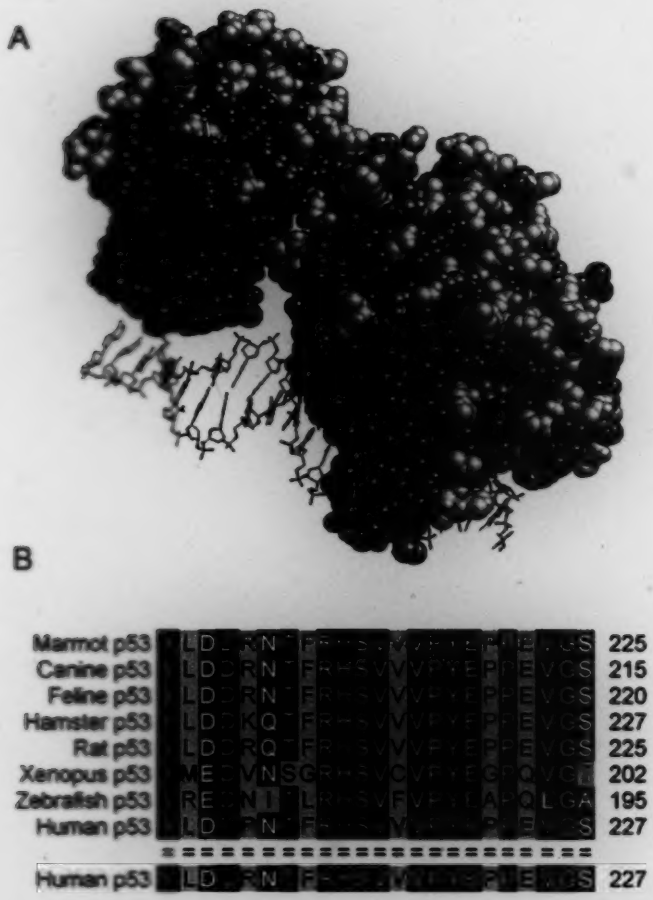

Figure 1. Sample output of COMBOSA3D showing (A) a colored three-dimensional representation of p53 (Cho et al. 1994) binding to DNA and (B) a portion of the colored reference alignment. Residues that are identical among the comparison sequences (the sequences located above the row of equal signs in B) are colored black, while similar residues are colored dark gray. Residues that are not identical or similar among 80% of the sequences are colored light gray.